

RA Pfaser C.M., Barrell B.;
 RT Genome sequence of the human malaria parasite *Plasmodium*
 RT *falciparum* ";
 RL Nature 419:498-511 (2002).
 DR EMBL; AB014841; AAN35938.1; -
 DR HSSP; 0931C2; 1MWT.
 KW Hypothetical protein.
 SQ SBOUENR 2227 AA; 262841 MW; 6D5DA3BE643339C CRC64;

Query Match 6.3%; Score 252.5; DB 2; Length 2227;
 Best Local Similarity 19.1%; Pred. No. 0.024;
 Matches 190; Conservative 155; Mismatches 319; Indels 329; Gaps 49;

```

QY 10 PRLANGKE-----GSLKDTTGVH-----HHQENESIKER 42
DB 690 YNNNNNNKCTCTSIKISENKKYPPFKSHVNSYKNTNHLPHRNALTSNNRNEEYER 749
QY 43 SPTIDRNIST-----IRDPENK---DLKLI-----KKKREVDPT-----SS 79
DB 750 EK---DRNITGNNNYLVYNNSCIPPLKQMIPIDGVRNKSINKLANNVTNQTSSVS 806
QY 80 TGRNBEVDY-----KYDKGNIIVADGTDLSYTEKLDKISKIYGVLSPSK 128
DB 807 TKNIDENSFDPPIINGIRSKYSNNNNI---NNSIGFNSKLDNYHQSMNV----- 858
QY 129 DGHPEILGKISVSKAKYVGNVYSIKIKATKVDHSHKMTF-----DLYANIN--- 179
DB 859 -----NESTPLKMMKKNNY-----IENYDDNNIFLVKNTEDTISNHNIGI 900
QY 180 -----DIVDGLAFAGDMR-----LPVKDN---DQKK-----DQKK----- 202
DB 901 HENSLKQVNLKACCTFHGYSRHHQNMVTTBEANLNQKNVSHYHNGTVLKLPLVNTNN 960
QY 203 -----AEIKIMPEKIKETKS---EYFVSSYGVNIEHGBGLSKOKPDNLTRMSGKI 253
DB 961 VAVNFPADINLSAQKRLSLKMGYEDKSMENYRNKIYNNINNNNNNNNDN-----NI 1013
QY 254 YSDSEKQOY---LKNONILIRKGYAL-----KVTYVPGKTDMLBGNVYSKEDIK--- 302
DB 1014 YNDNEYCQYNNSCYCFDHSDLKMMPLNHNQSKLTLTHSNKNSPFGNINVESKHLANPEI 1073
QY 303 -----IOKANP-----N-----N-----N-----N-----N 309
DB 1074 KTFANHSYPIINOGILNCPLQCLGYDSNQNRKNHVVYIKRVEYLNKNGISIIINVLKREG 1133
QY 310 LRAALS-----ETIYADSANVEDGRSTOSVLSALDGFNIIRYQVFTPKMD 356
DB 1134 LRKISTHNGKFPESFSNMDKRNVMYEGIANIQDNNNNNNKESCDNFKHRTKSLNFSRE 1192
QY 357 K-GE--AIDKQNLVYDSKLVLPFK-DDEKYYGEDKFPNVBAIKED-GSMLFTD---TKP 408
DB 1193 SYGHEKSLDVYQECYVKNKKLI--KRVNKKY--EDNNNSYLNEDNANMOPFBETNSNP 1248
QY 409 VNLSDM---KNYFNPSSKNTIYVRNPEYLRGKISDXGGFNWE-----LRVNESV----- 455
DB 1249 YIVDGENNNKKNVYV---NVLNNNSNYVYDSKDYDKSKENAKENKSDDIILNKENIHTLKD 1304
QY 456 -----VNNYLTYGDLHINTRDENIKLVKDDGDIWDW---GMKDYKANGPPDKV--TMDG 506
DB 1305 QCKKIQNNNEPFSBOADIENINNSOBEVYEKEHBP-LWYNASNEKKSYBELIYSDMS 1363
QY 507 NVYLTQYSDLNAKAVGVHYPFLYDNVVKPEVNIIDPKGNTSIEYADGKSVVFNINDKRNNG 566
DB 1364 NRVTKGKSDVANNVEFLANEDNLITJEKVKQLEKENKMIIMVETVBEININTIKTENTND 1423
QY 567 PDGEIQBQHIIYNGKEYSFPNDIKQIDITLNIKIYVDPART--TVK--EPIIANKNG 622
DB 1424 INEVR-----NEQRRSININD-----TWINHIIIDEPNDYPIKDIIECVHNNEN 1472
QY 623 BVSELKPHRVVTIIONGKEMSSSTIVSEBDPIIPVYKGELEKGYQPDGMEISGFEKXGAG 682
DB 1473 MYSNIBQYTFPHYDTRNNHVDK---NNQYFI-----FEEBGLANLNEBKK--V 1516

```

```

QY 683 YVINLSKD-----TPIKPVFKIEKKKEENKPTP-----DVSXGKON 720
DB 1517 YLENTYKDDHKGDGSKTSNLTSLNNTICSENDHNEKENTYVVRKGEKQIKKRVSHKQON 1576
QY 721 PQVN-HSQLN-----ESHKEDLQREBSOKS 746
DB 1577 EKLNEENYIINNYYDKMDNHRQNDITCKENDBEEN 1609

```

Search completed: February 16, 2005, 19:19:52
 Job time : 188 secs

This Page Blank (uspto)

LENGTH: 1787
 TYPE: PRT
 ORGANISM: Kl Parasite Clone
 US-10-415-253-2

Query March 5.0%; Score 200.5; DB 15; Length 1787;
 Best Local Similarity 20.5%; Pred. No. 0.015;
 Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

```

Qy 6 AES--KERN--LGNGESGLKDT-----TGVEHHQENB-----ESIX 40
Db 868 AESVTTSTNLIBEIQENTITNDTIEKLEBIHENVLSAALEVQSBKEKEVLDVIEYK 927
Qy 41 EKSSFTI-----DRNISTIRD-FEN-----KDLKGLKKKGRBYDDPTS 78
Db 928 EEVATTLLETVEQAEBKSANTITTEIFENLENNAVESNENNAENLEKLBETVFNVLDKVE 987
Qy 79 ET-----GKRMK--EYDYKY-----DDKG--NIIAYDDGTDLLEYET-----EKLD-- 114
Db 988 ETVEISGSLENNEMDKAFPFSEIFPNVKGIOENLIT---GMFRSLETSLVIOSEKEDLN 1044
Qy 115 -EIKSKIYGLSPSKDGHFIIKGISNVSKNAKV-----YGNNYKSIIEIKATYDPH 166
Db 1045 ENVVSSIIDNIEHMKEG---LLNKLKLENIISTEGVQETVEHVEQVYVDVDPAMK--- 1097
Qy 167 SKMTFPIYANINDVQGLAFAGDMRLFVKNDDKCAEIKIRMPKIKETKSEYVYSSY 226
Db 1098 -----DQFLGLINEAGL-----KEMFPNLEDFKSSDVTVEIEIKDEPVQKEVEKET 1146
Qy 227 GNVIELGRGD---LSKRPNPLTGM-----ESGKIYSDSEKQOYLKONITLRGYALX 277
Db 1147 VSIIIEEMENIVDVLEEKEDLTQMDAVESIISSDSKETESIKOX---EKDVSLV 1203
Qy 278 VTTNPG-----KTDMLEGNGVYSK--EDIAKIQGANPL----- 310
Db 1204 VEEVQDNMDESVEKYLELKNMBELMKDAVEINDITSLEIETQBLNEVEADLIRKMEK 1263
Qy 311 -----RALSETTIVADSRNVEGR--STOSVLSALDGFNIIRYQVFPKMDKGEALDK 363
Db 1264 LKELEKALSE---DSKEIIDAQDITLKVIEEHD-----ITTLDEVVELKDV 1309
Qy 364 DGNLVTDSSKLVLPKGDKREYTGBDKFVNEALKEDGSMFLDTKPEVNL-----SM 413
Db 1310 BBDKIKKXVDL---KOLBEDILKE---YKBIKELSEILIEDYKELKTIEITDILEKKKI 1362
Qy 414 DKATFNPESKNIYVNPPEPTLRGKISDXGCGFNMELRVNESVVD-----NYLIYGL 465
Db 1363 EKDHKEKEEBAEBEIKDLEADILKEVS-----SLEVEERKLEEVHELKEVEYHIIISGDA 1417
Qy 466 HINDTRDNIKLNVKD--GDIN-----DMGMKDYKANGPPDKVTMDGNVYLQTVGSDLN 518
Db 1418 HIKGLBEDDLR-EVDDLGKSLIDMLKGMELGDMKESLIEDVTTKGERV---ESLKDVL 1473
Qy 519 AKAVGVHYQFLYDN-----VXPEVNIIDPKGNTSIEYADGKSVENINDK--RN 564
Db 1474 SSALGMDREOKMTRKKAQRPKLEEVILKBEVKBPPKKIT-----KKKVRFDIKDEKPRD 1528
Qy 565 NGPDEHIOGHHIYNGK-----EYTSFNDIKQIIDKTILN-----IKIYVQDFAR--N 609
Db 1529 EIVVEYEMQDEIDEEVEBDIEDIEDVEDIDEDIDEDIGEDKDEVIDLIVQKERRIEK 1588
Qy 610 TTVKEPFIANKDTGE--VSELKPH-----RVTVTIONGKEMSTIIVSEBD 651
Db 1589 VRKKKKLKKYKEBEGVGLKGVDEVMKVQKIDKEVDKEVSKALESKNDVTNNVLKOND 1648
Qy 652 FILPYKGELEKGYQPDGWEISGFEGKCDAGYVNL-----SKDTFIKPVK 698
Db 1649 FFSKV-KAFVVKYKFAAFISAVAA--FASYVVGFTFSLSSCWTIIASSTYLLSKVDK 1705
Qy 699 KIBKKKEBENP---TFDV-----SKKDNPOVNSQLNESHKEDLQR 739
Db 1706 TINKKE---RPFYSFVFDIFNOLKHYLQOKMKESKEKNNNVIEVT--NKAKEKGNVOY 1760

```

Qy 740 EHSQKSDSTQVATVLDKNNISSKSTTNNPK 773
 Db 1761 TINKET-----TKVDKNNKVPRKRRTOQSK 1766

Search completed: February 16, 2005, 19:32:38
 Job time : 147 secs

This Page Blank (uspto)


```

RESULT 15
US-09-949-016-11433
; Sequence 11433, Application US/09949016
; Patent No. 681239
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PR
; ORGANISM: Human
US-09-949-016-11433

```

Query Match 4.6%; Score 185; DB 4; Length 2733;

Best Local Similarity 20.0%; Pred. No. 0.0032;

Matches 170; Conservative 150; Mismatches 338; Indels 194; Gaps 39;

```

QY 21 LKQDTGVEHHQENEB-SIKERSSPTIDRNISTIRDFENKDLKGLIKK--FREYVD 75
DB 1319 LKERIAGLEBEKQKNEFSQTLNENKYLTSQIST-KDGBIKMLQEBVTNNMLNQIOE 1377
QY 76 FTSEGRMEHYDYKVD-----KGNIIAY-DGDTDLSEYETKLBKISKIYG 122
DB 1378 ELSRVTLKTAEEKODLEBRLKQALALNGSIGNTQDVTDAQIKNELLESKMLKX 1437
QY 123 VLSPSKDHPEILCKISNVSQNAKYVY-----GN-----NYKSEIKXA 160
DB 1438 CVSELEBEKQOVLVEKTKVSEIRKREYLEKIQGAQKBERKNSHAKELQELKEKQOEVKQ 1497
QY 161 TKVD---FHSKTMFPDLXANINDIVDGLAFAGDKMLPYKNDQKKAEL-KIRMPKIKET 216
DB 1498 LQKDCIRYQEKISALERTVALE-----FVQTESQKULBETIKENLAQAVEHR 1544
QY 217 KSEPYVSYGVNVELEBGGDLSKNKPDMLTQWESGKIYSDSEKQOYLLKONIILK---- 272
DB 1545 KKAQAEIASFKVLLDDTQSEARVLADNLKGLKELQSKVESYKQMKQKQEDLERLEQA 1604
QY 273 -GYALKVTTYNPGKTDMLENGGVSKEDIACIQKAMPNLRALSTTIYADSRNVEDGERST 331
DB 1605 EEKLEKERNKQKQBLDMLRRKVLHRTIGRIQ-----VTLNKKQKEVQO---- 1649
QY 332 QSVLMSALDGNIRIYQVFTKMDKGEALIDKGLVTDSSKLVLPFGKDKKTYGDEKFN 391
DB 1650 ---LQENLDS--TVTQLAAFTSKMS---SLQDDRDVLTDAK---KMERKPSDAIQSK 1696
QY 392 VEA1--KEDG-SMLFIDTKPNLSMDKNYFNPSPK---SNKIY--VRNPEFIAGKISDK- 442
DB 1697 EEBIRLEKEDNCSYKQDLQKQSIMBELKINISRLNHOQIMBSKQCTEVOLQKXVCDTL 1756
QY 443 GGFNWEELRVNESVVDNLYIGDLHIIDTRFNKILNVKQGDIMQKMDQYANGFPDRTY 502
DB 1757 QGENKEILL-----SOLBRT-----HLTHSSQNELAKLESBIKSLKQDLT 1796
QY 503 DMD-----GNV---YLQTVSGL-NAKAVGVHVOFLYDNVPEV----- 537
DB 1797 DLSNLSKCKEQKQNLBGITIQOEDDIQNSK-----FSYBQLSTDLQASRELTSLAH 1849
QY 538 NIDPKGNTSIEYADGK--SVVFNINDKRNNGFDGRIQEOHIYINGKEYTSF---NDIKQI 592
DB 1850 EINNKEQKISILSGKBEALQVALAELRQO-HDKKIKELNMLSQEBEENIVLEBENKKA 1908

```

```

QY 593 IDKTLANIKIVKDFARNTTVASFILMKDTGVSBIK--PHEVTYTIQNGKMSSTIVABE 650
DB 1909 VDKTNQMLBETIKTKENIQQKQALDSFVKSMSLQNDRRIVGDYQQLBRRHSLILEK 1968
QY 651 DPLIPVTKGBLEKGYQPDGWEISGPEG-----KQDAGVINTLSQDT 691
DB 1969 DQLIQEAABENKCKE---EIRGLRSHMDIDINSBNKLDABLQYRBDLNVITTI-KDS 2023
QY 692 FIKQVEK-KIEKKKEENKPTFDVSKKQNDPOVNHQSOLNES-----HKKEDLQREBHSQK 745
DB 2024 QQKQULBVOQLQNNKLENTKAKLSEKLSSEBAN-EDLRSPNALQEBKQDLSKELBSIK 2082
QY 746 ---SDSTEDVTA 754
DB 2083 VSIQULTRQVTA 2094

```

Search completed: February 16, 2005, 19:20:41
Job time : 50 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comphen Ltd.

OM protein - protein search, using SW model

Run on: February 16, 2005, 19:06:53 ; Search time 170 Seconds
(without alignments)
1758.622 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGELAEKFKMLGNKGEGS.....ATVLDRNLSKSTTNPK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1808:*
2: geneseqp1908:*
3: geneseqp2008:*
4: geneseqp2008:*
5: geneseqp2008:*
6: geneseqp2003ae:*
7: geneseqp2003be:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	773	4 AAB48343	Aab48343 S. pneumo
2	4026	100.0	2140	6 ABU01020	Abu01020 S. pneumo
3	4026	100.0	2140	6 ABU45746	Abu45746 Protein e
4	4026	100.0	2140	8 ADM92113	Adm92113 S. pneumo
5	4015.5	99.7	2120	3 AAY81710	Aay81710 Streptoco
6	3789.5	94.1	2138	8 AdK48759	AdK48759 Streptoco
7	3048	75.7	637	8 ADR94534	Adr94534 Novel S.
8	750.5	18.6	1529	8 ADR96136	Adr96136 Novel S.
9	615	15.3	117	2 AAW5096	Aaw5096 Streptoco
10	615	15.3	117	5 ABP54590	Abp54590 S. pneumo
11	615	15.3	117	7 AdC45149	Adc45149 S. pneumo
12	227	5.6	1196	6 ABU24813	Abu24813 Protein e
13	222	5.5	861	7 ABO23608	AbO23608 Plasmodu
14	220	5.5	1166	8 AdK9186	AdK9186 Streptoco
15	218	5.4	1639	2 AAW54145	Aaw54145 P. falcip
16	218	5.4	1639	5 AAE29345	Aae29345 Plasmodu
17	217	5.4	1166	8 AdK9185	AdK9185 Streptoco
18	217	5.4	1166	8 AdK9178	AdK9178 Streptoco
19	216	5.4	1166	8 AdK9184	AdK9184 Streptoco
20	216	5.4	1166	8 AdK9183	AdK9183 Streptoco
21	216	5.4	1233	5 ABP30203	Abp30203 Streptoco
22	216	5.4	1233	5 ABP29675	Abp29675 Streptoco
23	216	5.4	1233	8 AdK9176	AdK9176 Streptoco
24	216	5.4	1233	8 AD010469	Ad010469 Group B S
25	216	5.4	1239	5 ABP25822	Abp25822 Streptoco

26	215.5	5.4	1254	2 AAR07503	Aar07503 Merozoite
27	215.5	5.4	1254	2 AAW24575	Aaw24575 Merozoite
28	215	5.3	1166	8 AdK9179	AdK9179 Streptoco
29	212.5	5.3	2485	3 AAB18172	Aab18172 Plasmodu
30	212	5.3	1141	6 ABU42327	Abu42327 Protein e
31	212	5.3	1188	3 AAB18183	Aab18183 Plasmodu
32	211.5	5.3	991	8 ADP86452	Adp86452 Clostridi
33	211.5	5.3	1104	7 ADG73658	Adg73658 C. perfr
34	209	5.2	1558	3 AAB18324	Aab18324 Plasmodu
35	208.5	5.2	3029	8 ADP25433	Adp25433 Plasmodu
36	207.5	5.2	4688	6 ABU48941	Abu48941 Protein e
37	207	5.1	1191	6 ABU24124	Abu24124 Protein e
38	207	5.1	1654	1 AAP50777	Aap50777 Sequence
39	207	5.1	1979	3 AAB18171	Aab18171 Plasmodu
40	205	5.1	2024	8 ADP25444	Adp25444 Plasmodu
41	204.5	5.1	1516	3 AAB18195	Aab18195 Plasmodu
42	204.5	5.1	2500	3 AAB18272	Aab18272 Plasmodu
43	204	5.1	1166	6 ABJ18982	Abj18982 Pathogen
44	204	5.1	1245	3 AAB18244	Aab18244 Plasmodu
45	203	5.0	1166	2 AAY08643	Aay08643 S. aureus

ALIGNMENTS

RESULT 1	AAB48343
ID	AAB48343 standard; protein; 773 AA.
AC	AAB48343;
DT	20-APR-2001 (first entry)
XX	XX
DE	S. pneumoniae Sp130 polypeptide.
XX	XX
KW	Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;
KW	bronchial; lung; blood; infection; immune response; immunotherapy;
KW	antibacterial; auditory; vaccine.
XX	XX
OS	Streptococcus pneumoniae.
XX	XX
PN	WO200076540-A2.
PD	21-DEC-2000.
XX	XX
PF	09-JUN-2000; 2000MO-US015925.
XX	XX
PR	10-JUN-1999; 99US-0138453P.
XX	XX
PA	(MEDI-) MED IMMUNE INC.
XX	XX
PI	Adamou JE, Choi GH;
XX	XX
DR	WPI; 2001-112197/12.
XX	XX
DR	N-PSDB; AAC84742.
XX	XX
PT	New vaccines comprising Sp128 or Sp130 polypeptides, for treating and
PT	preventing pneumococcal infections, particularly infections caused by
PT	Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
XX	blood infections.
XX	XX
PS	Claim 8; Page 51-54; 54pp; English.
XX	XX
CC	The invention relates to novel immunogenic polypeptides, Sp128 and Sp130
CC	from S. pneumoniae. Vaccines comprising the polypeptides are useful for
CC	the treatment and prevention of pneumococcal infections, particularly
CC	infections caused by Streptococcus, such as otitis media, nasopharyngeal,
CC	bronchial, lung or blood infections. The antigens are used as immunogeni
CC	c agents to stimulate an immune response. The antisera and antibodies may
CC	also be used in diagnosing and treating pneumococcal infections.
CC	Recombinant polypeptides serve as a mechanism for stimulating production
CC	of antibodies for use in passive immunotherapy, diagnostic reagents, and
CC	as reagents in other processes such as affinity chromatography. The

CC present sequence represents the *S. pneumoniae* Sp130 polypeptide
XX Sequence 773 AA;
SQ

Query Match 100.0%; Score 4026; DB 4; Length 773;
Best Local Similarity 100.0%; Pred. No. 7.5e-225;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGELAESKPKNLGNGKESGLKKDTTGVHHQENESIKESSTIIRNISTIRDPENK 60
DB 1 KLGELAESKPKNLGNGKESGLKKDTTGVHHQENESIKESSTIIRNISTIRDPENK 60
QY 61 DLKKLKKKKFREVDFTSETGKRMEEYDYKDDKNNIIAYDGTDLLEYETELDKSKSI 120
DB 61 DLKKLKKKKFREVDFTSETGKRMEEYDYKDDKNNIIAYDGTDLLEYETELDKSKSI 120
QY 121 YGVLSPSKQGHFEILLKISNVSKNAKYVGNKYKSIKATKDPFSKMTFDLYANIND 180
DB 121 YGVLSPSKQGHFEILLKISNVSKNAKYVGNKYKSIKATKDPFSKMTFDLYANIND 180
QY 181 IYDGLAPAGDMRLFVKNDQKAEIKIRMEPIKETSEYVSYGVNIELEGDDLKSN 240
DB 181 IYDGLAPAGDMRLFVKNDQKAEIKIRMEPIKETSEYVSYGVNIELEGDDLKSN 240
QY 241 KPDNLTGMESGKLYSDSEKQOYLKNNIILRKGYALKVTTYPNGKTDMEGNGVSKEDI 300
DB 241 KPDNLTGMESGKLYSDSEKQOYLKNNIILRKGYALKVTTYPNGKTDMEGNGVSKEDI 300
QY 301 AKIQKANPMLRALSETTIYADSHNVEDGRSTOSVLSALDGFNIIYYOFTFPMNDKGA 360
DB 301 AKIQKANPMLRALSETTIYADSHNVEDGRSTOSVLSALDGFNIIYYOFTFPMNDKGA 360
QY 361 IDKQGNLVTDSSKLVFGKODKEYTGEDKFNVAIKEDGSMLEFDITKPVNLSMDKNYFNP 420
DB 361 IDKQGNLVTDSSKLVFGKODKEYTGEDKFNVAIKEDGSMLEFDITKPVNLSMDKNYFNP 420
QY 421 SSKNKTIVNRPFFYLRGKISDKGFWMELRVNESVVDNVLIGDLHIDNTRDPIKLNVK 480
DB 421 SSKNKTIVNRPFFYLRGKISDKGFWMELRVNESVVDNVLIGDLHIDNTRDPIKLNVK 480
QY 481 DGDIMDMGKMDYKANGFPDKVTMDGNVYLQTSYSDLNAAVGVHGYQFLYDNVKKPEVND 540
DB 481 DGDIMDMGKMDYKANGFPDKVTMDGNVYLQTSYSDLNAAVGVHGYQFLYDNVKKPEVND 540
QY 541 PKGNNTSIEYADGKSVFNNIDKRNNGPDEIOEHYIYNGKEYTSFNDIKQIIDKTLNKK 600
DB 541 PKGNNTSIEYADGKSVFNNIDKRNNGPDEIOEHYIYNGKEYTSFNDIKQIIDKTLNKK 600
QY 601 IYVKDPAARTTYKEFLINKDTGVSFLKXPHRYVTVTQNGKEMSTTVSEDFILPYKGE 660
DB 601 IYVKDPAARTTYKEFLINKDTGVSFLKXPHRYVTVTQNGKEMSTTVSEDFILPYKGE 660
QY 661 LKRGYQFDGMEISGFEGKDGAGVNLISKDTFLKPVFKKLEKCEENKPTFVDSKKON 720
DB 661 LKRGYQFDGMEISGFEGKDGAGVNLISKDTFLKPVFKKLEKCEENKPTFVDSKKON 720
QY 721 PNVNHSQNLNESHKEDLQREHSQKSDSTKDVATVATVLDKNNISSKSTNNPNK 773
DB 721 PNVNHSQNLNESHKEDLQREHSQKSDSTKDVATVATVLDKNNISSKSTNNPNK 773

RESULT 2
ABU01020
ID ABU01020 standard; protein; 2140 AA.
XX
XX
AC ABU01020;
XX
XX
DT 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
XX
DE *S. pneumoniae* type 4 strain protein from coding region #590.
XX
XX
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;

KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX
XX OS Streptococcus pneumoniae; type 4 strain.
XX
XX PN WO200277021-A2.
XX
XX PD 03-OCT-2002.
XX
XX PF 27-MAR-2002; 2002WO-1B002163.
XX
XX PR 27-MAR-2001; 2001GB-00007658.
XX
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX
XX PI Masignani V, Tettelin H, Fraser C;
XX
XX DR WPI, 2003-040579/03.
XX
XX DR N-PSDB; ABX06302.
XX
XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
XX PT useful as medicaments for treating or preventing a disease or infection
XX PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
XX PT ear infection.
XX
XX PS Claim 1; SEQ ID NO 1180; 56pp; English.
XX
XX
XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,
XX expressed from 2469 of 2489 identified DNA coding regions from the
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX AB556454. Also included are an antibody which binds one of the proteins,
XX treating a patient by administering the protein, DNA or antibody (in a
XX composition), a kit comprising first and second primers, which are the
XX nucleic acid cited above or fragments between nucleotides 8-100 of a
XX sequence not defined in the specification, for amplifying a target
XX sequence contained within a Streptococcus nucleic acid sequence, where
XX the first primer is substantially complementary to the target sequence
XX and the second primer is substantially complementary to the complement of
XX the target sequence, and where the parts of the primers having
XX substantial complementarity define the termini of the target sequence to
XX be amplified, assay comprising contacting a test compound with the
XX protein, and determining whether the test compound binds to the protein
XX and a Streptococcus pneumoniae bacterium, where one or more genes
XX encoding the protein has been rendered inactive. The proteins, nucleic
XX acid molecules, antibody and compositions are useful as medicaments for
XX treating or preventing a disease or infection due to streptococcus
XX bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis
XX media or ear infection. They are also useful in developing vaccines,
XX diagnostic and antibiotics. The methods are useful for identifying
XX immunodominant proteins. The present sequence is one of the 2469 proteins
XX expressed by the identified coding regions from the genomic sequence.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
XX standardise OS field)
XX
XX SQ Sequence 2140 AA;
XX

Query Match 100.0%; Score 4026; DB 6; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.9e-224;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGELAESKPKNLGNGKESGLKKDTTGVHHQENESIKESSTIIRNISTIRDPENK 60
DB 1334 KLGELAESKPKNLGNGKESGLKKDTTGVHHQENESIKESSTIIRNISTIRDPENK 1393
QY 61 DLKKLKKKKFREVDFTSETGKRMEEYDYKDDKNNIIAYDGTDLLEYETELDKSKSI 120
DB 1394 DLKKLKKKKFREVDFTSETGKRMEEYDYKDDKNNIIAYDGTDLLEYETELDKSKSI 1453

QY 121 YGVLSPSKDGHEFLIGKISNVSNAKAVYGNMYKSIIEIKATKYDFHSTMTFEDLVANIND 180
 DB 1454 YGVLSPSKDGHEFLIGKISNVSNAKAVYGNMYKSIIEIKATKYDFHSTMTFEDLVANIND 1513
 QY 181 IYDGLAFAGDMRLFYVNDNDQKAEIKIRMEPKIKETKSEYPVYSSYGNVIEIGEGDLSKN 240
 DB 1514 IYDGLAFAGDMRLFYVNDNDQKAEIKIRMEPKIKETKSEYPVYSSYGNVIEIGEGDLSKN 1573
 QY 241 KPDNLTKMESGKIYSESEKQOYLKONILIRKGYALKVTTYNNGKTDMEGNGVSKEDI 300
 DB 1574 KPDNLTKMESGKIYSESEKQOYLKONILIRKGYALKVTTYNNGKTDMEGNGVSKEDI 1633
 QY 301 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGNRIIRYOVFTFRANDKGBA 360
 DB 1634 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGNRIIRYOVFTFRANDKGBA 1693
 QY 361 IYDGLAFAGDMRLFYVNDNDQKAEIKIRMEPKIKETKSEYPVYSSYGNVIEIGEGDLSKN 420
 DB 1694 IYDGLAFAGDMRLFYVNDNDQKAEIKIRMEPKIKETKSEYPVYSSYGNVIEIGEGDLSKN 1753
 QY 421 SKSNKIYVNPPEFYLGKISDKGFMWELRVNSVVDNYLIYDGLHIDNTRDENIKLANVK 480
 DB 1754 SKSNKIYVNPPEFYLGKISDKGFMWELRVNSVVDNYLIYDGLHIDNTRDENIKLANVK 1813
 QY 481 DGDIMDMGKDYKANGFPDKVYTDMDGNVYLQTSYSDLNKAVGVHYQFLVDNYPKPEVND 540
 DB 1814 DGDIMDMGKDYKANGFPDKVYTDMDGNVYLQTSYSDLNKAVGVHYQFLVDNYPKPEVND 1873
 QY 541 PKGNISIEIADGKSVFVFNINDKRNNGPDGEIOGHYIYNGKEYTSRNDIKQIIDKTLNLIK 600
 DB 1874 PKGNISIEIADGKSVFVFNINDKRNNGPDGEIOGHYIYNGKEYTSRNDIKQIIDKTLNLIK 1933
 QY 601 IYVVDARATTVYEFILNKDTGSEVSELKPHRVVVTIIONGKEMSTIYSEDFILPYKGE 660
 DB 1934 IYVVDARATTVYEFILNKDTGSEVSELKPHRVVVTIIONGKEMSTIYSEDFILPYKGE 1993
 QY 661 LEKGYPDGMELISGFEKGDAGVILNLSKDTFIKPVFKKLEKKEBENKTFPDVSKKON 720
 DB 1994 LEKGYPDGMELISGFEKGDAGVILNLSKDTFIKPVFKKLEKKEBENKTFPDVSKKON 2053
 QY 721 PQVNHQALMESHRKEDLOREBHSQKSDSTVDATVATLADKNISKSSTNNPNK 773
 DB 2054 PQVNHQALMESHRKEDLOREBHSQKSDSTVDATVATLADKNISKSSTNNPNK 2106
 RESULT 3
 ABU45746
 ID ABU45746 standard; protein; 2140 AA.
 AC ABU45746;
 XX
 DT 19-JUN-2003 (first entry)
 DE Protein encoded by Prokaryotic essential gene #31273.
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.
 OS Streptococcus pneumoniae.
 XX
 PN WO200277183-A2.
 PD 03-OCT-2002.
 PF 21-MAR-2002; 2002WO-US009107.
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyekind JW,
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA49616.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation or
 PT isolate candidate molecules for rational drug discovery programs.
 PS Claim 25; SEQ ID NO 73670; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 623 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SO Sequence 2140 AA;
 Query Match 100.0%; Score 4026; DB 6; Length 2140;
 Best Local Similarity 100.0%; Pred. No. 2.9e-224;
 Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KLGELAESKFKNLGNGKESLKKDITGVHHOENESIKESSPFIDNISTIRPFENK 60
 DB 1334 KLGELAESKFKNLGNGKESLKKDITGVHHOENESIKESSPFIDNISTIRPFENK 1393
 QY 61 DLKLIKKKFRVDDFTSETGRMEBYDYKDDKGNIIAYDDGTDLEVEKELDEIKSKI 120
 DB 1394 DLKLIKKKFRVDDFTSETGRMEBYDYKDDKGNIIAYDDGTDLEVEKELDEIKSKI 1453
 QY 121 YGVLSPSKDGHEFLIGKISNVSNAKAVYGNMYKSIIEIKATKYDFHSTMTFEDLVANIND 180
 DB 1454 YGVLSPSKDGHEFLIGKISNVSNAKAVYGNMYKSIIEIKATKYDFHSTMTFEDLVANIND 1513
 QY 181 IYDGLAFAGDMRLFYVNDNDQKAEIKIRMEPKIKETKSEYPVYSSYGNVIEIGEGDLSKN 240
 DB 1514 IYDGLAFAGDMRLFYVNDNDQKAEIKIRMEPKIKETKSEYPVYSSYGNVIEIGEGDLSKN 1573
 QY 241 KPDNLTKMESGKIYSESEKQOYLKONILIRKGYALKVTTYNNGKTDMEGNGVSKEDI 300
 DB 1574 KPDNLTKMESGKIYSESEKQOYLKONILIRKGYALKVTTYNNGKTDMEGNGVSKEDI 1633
 QY 301 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGNRIIRYOVFTFRANDKGBA 360
 DB 1634 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGNRIIRYOVFTFRANDKGBA 1693

```
QY 361 IDKGNLYTDSKLVLFQKDKKEYTGSDKFNVEAIEDGSMLEFIDTKPVNLSDMKYFNP 420
DB 1694 IDKGNLYTDSKLVLFQKDKKEYTGSDKFNVEAIEDGSMLEFIDTKPVNLSDMKYFNP 1753
QY 421 SKSNKIYVRNPEFYLRGKISDKGCFNMELRVNESVVDNYLYGDLHINDTRDPNKLANK 480
DB 1754 SKSNKIYVRNPEFYLRGKISDKGCFNMELRVNESVVDNYLYGDLHINDTRDPNKLANK 1813
QY 481 DGDIMDMGKDYKANGFPDXTDMDGNVYLQOTGYSDLNKAVGVHYQPLVDNVKPEVND 540
DB 1814 DGDIMDMGKDYKANGFPDXTDMDGNVYLQOTGYSDLNKAVGVHYQPLVDNVKPEVND 1873
QY 541 PKGNTSIEYADGKSVFVFNINDKRNNGPDEIOEQHIYINGKEYTSFNDIKQIIDTLANK 600
DB 1874 PKGNTSIEYADGKSVFVFNINDKRNNGPDEIOEQHIYINGKEYTSFNDIKQIIDTLANK 1933
QY 601 IYVKDFARNTTYKEFLINKDTGSEVELKPHRYVTIIONGKEMSTIVSEEDFILPVYKGE 660
DB 1934 IYVKDFARNTTYKEFLINKDTGSEVELKPHRYVTIIONGKEMSTIVSEEDFILPVYKGE 1993
QY 661 LEKGYQFDGWEISGFEKKDAGVYVNLSDKDTFIKPVFKKIEBKKEENKPTFDVSKKDN 720
DB 1994 LEKGYQFDGWEISGFEKKDAGVYVNLSDKDTFIKPVFKKIEBKKEENKPTFDVSKKDN 2053
QY 721 POWNHSQLNESHKEDLQREHSHQKSDSTKYATATVLDKNNISSKSTNNPNK 773
DB 2054 POWNHSQLNESHKEDLQREHSHQKSDSTKYATATVLDKNNISSKSTNNPNK 2106

RESULT 4
ADM92113
ID ADM92113 standard; protein; 2140 AA.
AC ADM92113;
XX
XX 03-JUN-2004 (first entry)
DE S pneumoniae antigenic protein sequence SegidJ10.
XX
XX antibacterial; gene therapy; Streptococcus pneumoniae infection;
XX antigenic.
OS Streptococcus pneumoniae.
XX
XX WO2004020609-A2.
XX
XX 11-MAR-2004.
XX
XX PD 02-SEP-2003; 2003WO-US027401.
XX
XX PR 30-AUG-2002; 2002US-0407082P.
XX
XX PA (TUFT ) UNIV TUFTS.
XX
XX PI Cam1111 A, Hava DL;
XX
XX DR WPI; 2004-239189/22.
XX
XX DR N-PSDB; ADM91876.
XX
XX PT New Streptococcus pneumoniae nucleic acid molecules, useful for
XX PT diagnosing, treating and preventing active infections of Streptococcus
XX PT pneumoniae.
PS Claim 27; SEQ ID NO 310; 123bp; English.
XX
XX CC This invention relates to novel isolated Streptococcus pneumoniae nucleic
XX CC acid molecules and the antigenic polypeptides encoded by them. The
XX CC invention may be useful for the production of compounds with an
XX CC antibacterial activity or for gene therapy. The nucleic acid molecules,
XX CC compositions and methods disclosed are useful for treating Streptococcus
XX CC pneumoniae infection. The present sequence is that of an S pneumoniae
XX CC protein of the invention.
```

```
XX SQ Sequence 2140 AA:
Query Match 100.0%; Score 4026; DB 8; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2,9e-224;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KGEIAESKFKVLNGKSGSLKKDTTGVHHHQQNEESIKESFTIDRNISTIDPENK 60
DB 1334 KGEIAESKFKVLNGKSGSLKKDTTGVHHHQQNEESIKESFTIDRNISTIDPENK 1393
QY 61 DLKIKIKKKFRVDDFTSETRKMEBYDYKDDKNNIIAYDDGTLLEYTEKLDLISKI 120
DB 1394 DLKIKIKKKFRVDDFTSETRKMEBYDYKDDKNNIIAYDDGTLLEYTEKLDLISKI 1453
QY 121 YGVLSPKDGHFEILGKISNYSKNAKYVYGNVYSIEIKATKYDPSKTMFEDLYANIND 180
DB 1454 YGVLSPKDGHFEILGKISNYSKNAKYVYGNVYSIEIKATKYDPSKTMFEDLYANIND 1513
QY 181 IYDGLAFAGDMRLFYKNDQKKAELIKIMPEKIKETKSEYPVSSYGVNIEIGEDLSKN 240
DB 1514 IYDGLAFAGDMRLFYKNDQKKAELIKIMPEKIKETKSEYPVSSYGVNIEIGEDLSKN 1573
QY 241 KPDNLTNMSGKIVSDSEKQOYLKONTILIRGVALKTTYNPKTMDLEGNVYSKEDI 300
DB 1574 KPDNLTNMSGKIVSDSEKQOYLKONTILIRGVALKTTYNPKTMDLEGNVYSKEDI 1633
QY 301 AKIQANPLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYOVFTFKANDKGBA 360
DB 1634 AKIQANPLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYOVFTFKANDKGBA 1693
QY 361 IDKGNLYTDSKLVLFQKDKKEYTGSDKFNVEAIEDGSMLEFIDTKPVNLSDMKYFNP 420
DB 1694 IDKGNLYTDSKLVLFQKDKKEYTGSDKFNVEAIEDGSMLEFIDTKPVNLSDMKYFNP 1753
QY 421 SKSNKIYVRNPEFYLRGKISDKGCFNMELRVNESVVDNYLYGDLHINDTRDPNKLANK 480
DB 1754 SKSNKIYVRNPEFYLRGKISDKGCFNMELRVNESVVDNYLYGDLHINDTRDPNKLANK 1813
QY 481 DGDIMDMGKDYKANGFPDXTDMDGNVYLQOTGYSDLNKAVGVHYQPLVDNVKPEVND 540
DB 1814 DGDIMDMGKDYKANGFPDXTDMDGNVYLQOTGYSDLNKAVGVHYQPLVDNVKPEVND 1873
QY 541 PKGNTSIEYADGKSVFVFNINDKRNNGPDEIOEQHIYINGKEYTSFNDIKQIIDTLANK 600
DB 1874 PKGNTSIEYADGKSVFVFNINDKRNNGPDEIOEQHIYINGKEYTSFNDIKQIIDTLANK 1933
QY 601 IYVKDFARNTTYKEFLINKDTGSEVELKPHRYVTIIONGKEMSTIVSEEDFILPVYKGE 660
DB 1934 IYVKDFARNTTYKEFLINKDTGSEVELKPHRYVTIIONGKEMSTIVSEEDFILPVYKGE 1993
QY 661 LEKGYQFDGWEISGFEKKDAGVYVNLSDKDTFIKPVFKKIEBKKEENKPTFDVSKKDN 720
DB 1994 LEKGYQFDGWEISGFEKKDAGVYVNLSDKDTFIKPVFKKIEBKKEENKPTFDVSKKDN 2053
QY 721 POWNHSQLNESHKEDLQREHSHQKSDSTKYATATVLDKNNISSKSTNNPNK 773
DB 2054 POWNHSQLNESHKEDLQREHSHQKSDSTKYATATVLDKNNISSKSTNNPNK 2106

RESULT 5
AAV81710
ID AAV81710 standard; protein; 2120 AA.
AC AAV81710;
XX
XX 02-JUN-2000 (first entry)
DE Streptococcus pneumoniae protein sequence ID3.
XX
XX DR Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
XX KW bacterial pneumonia; aplenia; heart disease; lung disease; alcoholism;
XX kidney disease; diabetes; immunosuppressive disorder; otitis media;
```

KW pneumococcal septicemia; sinusitis; meningitis; therapy.
 XX Streptococcus pneumoniae.
 XX WO200006738-A2.
 XX PN
 PD 10-FEB-2000.
 XX PF 27-JUL-1999; 99WO-GB002452.
 XX PR 27-JUL-1998; 98GB-00016336.
 XX PR 19-MAR-1999; 99US-0123529P.
 XX PA (MICR-) MICROBIAL TECHNIKS LTD.
 PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
 XX MPI, 2000-195301/17.
 DR N-PSDB; AA291806.
 PT Streptococcal proteins and polynucleotides useful for diagnosis,
 PT treatment and prophylaxis of bacterial infections.
 XX
 XX Claim 2, Page 41-42; 76pp; English.
 PS This sequence represents a Streptococcus pneumoniae protein of the
 CC invention. The proteins (or their homologues, derivatives and/or
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
 CC compositions comprising the proteins are useful as vaccines and also in
 CC diagnostic assays. The sequences are useful for the detection or
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
 CC with them. Agents capable of antagonising, inhibiting or interfering with
 CC the function or expression of the protein or polypeptide are useful in
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae
 CC infection. As the sequences can be used to treat S. pneumoniae infection,
 CC they can be used to treat bacterial pneumonia, which has high rates in
 CC young children, the elderly, and in patients with predisposing conditions
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
 CC or with immunosuppressive disorders, especially AIDS. They can also be
 CC used to treat pneumococcal septicemia, otitis media, sinusitis, and
 CC meningitis.
 CC
 SQ Sequence 2120 AA;
 Query Match 99.7%; Score 4015.5; DB 3; Length 2120;
 Best Local Similarity 99.9%; Pred. No. 1.1e-223;
 Matches 773; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 KLGEIAESKRNKNGKESGLKKDTTGVHHQENESIEKESFTIDRNISTRDPEFK 60
 DB 1313 KLGEIAESKRNKNGKESGLKKDTTGVHHQENESIEKESFTIDRNISTRDPEFK 1372
 QY 61 DLKLLIKKKFREYDDTSETGKMEYDYKYDKGNIIAVDGTDLLEYETKDEIKSKI 120
 DB 1373 DLKLLIKKKFREYDDTSETGKMEYDYKYDKGNIIAVDGTDLLEYETKDEIKSKI 1432
 QY 121 YGVLSPSKOGHEIILKISVNSKNAKYVYGNMYKSIIEKATKYDPHKTMTPLVANNID 180
 DB 1433 YGVLSPSKOGHEIILKISVNSKNAKYVYGNMYKSIIEKATKYDPHKTMTPLVANNID 1492
 QY 181 YVDGLAFAGDMRLFVKDNDQKAEIKIRMEKIKETKSEYVYSSYGVNIELEGGDLISKR 240
 DB 1493 YVDGLAFAGDMRLFVKDNDQKAEIKIRMEKIKETKSEYVYSSYGVNIELEGGDLISKR 1552
 QY 241 KPNULTKMSGKTIYSSEKQYILKONIIIRKGYALKVTTYNPKTDMLEGNVYSKEDI 300
 DB 1553 KPNULTKMSGKTIYSSEKQYILKONIIIRKGYALKVTTYNPKTDMLEGNVYSKEDI 1612
 QY 301 AKIQKANPILRALSETTIYADSRNVEDGSGTOSVLSALDPFYIIRYQVTFQANDKGEA 360
 DB 1613 AKIQKANPILRALSETTIYADSRNVEDGSGTOSVLSALDPFYIIRYQVTFQANDKGEA 1672
 QY 361 IDKGNLVTDSKLVLFKGDKEKYEYTGEDKFNVAIKEDSGMLPIDTRPVNLSDMKVYFNP 420

DB 1673 IDKGNLVTDSKLVLFKGDKEKYEYTGEDKFNVAIKEDSGMLPIDTRPVNLSDMKVYFNP 1732
 QY 421 SKSNKIYVNPPEFYLRGKISDKGFWMLRVNESVVDNVLIGDLHIDNTR-DEFNKLVN 479
 DB 1733 SKSNKIYVNPPEFYLRGKISDKGFWMLRVNESVVDNVLIGDLHIDNTRDPEFKLVN 1792
 QY 480 KGDIDMDGMDKYKANGFPDKYTDMDGANVYLTQTSGLNAXAVGVYQFLYDNVKEVNI 539
 DB 1793 KGDIDMDGMDKYKANGFPDKYTDMDGANVYLTQTSGLNAXAVGVYQFLYDNVKEVNI 1852
 QY 540 DPKGNTSIYADGKSVFENINDKRNNGFDEIIOEHIYINGKYSFNDIKOIIDKTLNI 599
 DB 1853 DPKGNTSIYADGKSVFENINDKRNNGFDEIIOEHIYINGKYSFNDIKOIIDKTLNI 1912
 QY 600 KIVVDFPANTVKEIILNKDGEVSELKPHVVTYITQNGKMSSTVSEEDFILPVYNG 659
 DB 1913 KIVVDFPANTVKEIILNKDGEVSELKPHVVTYITQNGKMSSTVSEEDFILPVYNG 1972
 QY 660 ELEKGYQPDGWEISGFEKGDAGVYINLSKDTFIKPVFKIEEKEENKPTFDVSKKD 719
 DB 1973 ELEKGYQPDGWEISGFEKGDAGVYINLSKDTFIKPVFKIEEKEENKPTFDVSKKD 2032
 QY 720 NQVNHSQLNESHKREDLQREHSHQKSDSTKQVATVLDKNNISSKSTNNPNK 773
 DB 2033 NQVNHSQLNESHKREDLQREHSHQKSDSTKQVATVLDKNNISSKSTNNPNK 2086
 RESULT 6
 ADK48759
 ID ADK48759 standard; protein; 2138 AA.
 AC ADK48759;
 XX 20-MAY-2004 (first entry)
 DT Streptococcus pneumoniae protein, Seq ID No 5274.
 DE Streptococcus pneumoniae protein, Seq ID No 5274.
 XX Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
 XX Streptococcus pneumoniae.
 OS US6699703-B1.
 PN 02-MAR-2004.
 PD 26-MAY-2000; 2000US-00583110.
 PF 02-JUL-1997; 97US-0051553P.
 PR 12-MAY-1998; 98US-0085131P.
 PR 30-JUN-1998; 98US-00107433.
 XX (GENO-) GENOME THERAPEUTICS CORP.
 PA Doucette-Stamm J, Bush D, Zeng Q, Opperman T, Houseweart CE;
 PI MPI: 2004-212399/20.
 DR N-PSDB; ADK46098.
 XX New nucleic acid molecules and polypeptides useful for diagnosing,
 PT preventing and treating pathological conditions resulting from bacterial
 PT infection, e.g. Streptococcus pneumoniae infection, and in drug
 PT screening.
 XX Disclosure, SEQ ID NO 5274; 301pp; English.
 PS The invention relates to isolated Streptococcus pneumoniae nucleic acids
 XX and polypeptides. The nucleic acids and proteins are useful for
 CC diagnosing, preventing and treating pathological conditions resulting
 CC from bacterial infection, such as S. pneumoniae infection. These may also
 CC be used for drug screening procedures. The present sequence represents a
 CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
 CC data for this patent did not appear in the printed specification but was

CC obtained in electronic format directly from USPRO at
CC segdata.uspto.gov/sequence.html.

XX Sequence 2138 AA;

Query Match 94.1%; Score 3789.5; DB 8; Length 2138;
Best Local Similarity 94.8%; Pred. No. 1.5e-210;
Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;

```

QY 1 KLGEIAESKPKNLGNGKESLKKDDTGVHHHQQEENESIKESKSPFTIDNISTIDFENK 60
DB 1333 KLGEIPESKPKNLKVKKDDSLKKEPTEVNNLLVNDQSLKESLKFNIHITIDFENK 1392
QY 61 DLKLLIKKKFREVDFSTSTGKMEYDYKTDKGNILAYDDGTDLLEYTEKLDKSKKI 120
DB 1393 DLKLLIKKKYKQEDDFVN-CGTRIVRDYKDDKGNILAYDDGTDLLEYTEKLDKSKKI 1451
QY 121 YGVLSPSKDGHEILGKISNVSKNAKVYGNVYKSIKATKYDHFPSKMTEDLVANIND 180
DB 1452 YGVLSPSKDGHEILGKISNVSKNAKVYGNVYKSIKATKYDHFPSKMTEDLVANIND 1511
QY 181 IVDGLAFAGDMELFVNDNDOKTAETIKIRMPKIKETSEYVYSSYGVNIELEGEDLSKN 240
DB 1512 IVDGLAFAGDMELFVNDNDOKTAETIKIRMPKIKETSEYVYSSYGVNIELEGEDLSKN 1571
QY 241 KEDNLTWESGKIYSDSEKQYLLKDNILIRKGVALKVTYTPGKTDMLENGVYSKEDI 300
DB 1572 KEDNLTWESGKIYSDSEKQYLLKDNILIRKGVALKVTYTPGKTDMLENGVYSKEDI 1631
QY 301 AKIQKAPLRLALSETTYIADSRNVEDGSTOSVMSALDGFNIIRYOVFTFRKMDKGBA 360
DB 1632 AKIQKAPLRLALSETTYIADSRNVEDGSTOSVMSALDGFNIIRYOVFTFRKMDKGBA 1691
QY 361 IDKDGMLVTSSKLVLFQGDKEKYEKEDKFNVEAIKEDGSMLEFDTKPVNLSMDKNYFNP 420
DB 1692 IDKDGMLVTSSKLVLFQGDKEKYEKEDKFNVEAIKEDGSMLEFDTKPVNLSMDKNYFNP 1751
QY 421 SSKNKIYVNNPEFYLRGKISDVGKGFVMELRVNESVVDNYLIIGDLHIDTRFNIKLVNK 480
DB 1752 SSKNKIYVNNPEFYLRGKISDVGKGFVMELRVNESVVDNYLIIGDLHIDTRFNIKLVNK 1811
QY 481 DDDIMDMGKDYKANGFPDKVTMDMGVNYLQTYSPDLNKAAGVHYQFLYDVKDEVND 540
DB 1812 DDDIMDMGKDYKANGFPDKVTMDMGVNYLQTYSPDLNKAAGVHYQFLYDVKDEVND 1871
QY 541 PKGNTSIEYADGKSVFENIDNRKNGFDEIOEHYIYNGKETSFNKQIIDKTLNIX 600
DB 1872 PKGNTSIEYADGKSVFENIDNRKNGFDEIOEHYIYNGKETSFNKQIIDKTLNIX 1931
QY 601 IYVQFAPARTTYKEFLANKDTGVESELKPHRYVTYTIQNGKEMSTTVSEEDTILPYKGE 660
DB 1932 IYVQFAPARTTYKEFLANKDTGVESELKPHRYVTYTIQNGKEMSTTVSEEDTILPYKGE 1991
QY 661 LEKGYOFDMEISGFEKGDAGVYINLSKDTFIKPVFKKIEEKKEENKPTDVSKKON 720
DB 1992 LEKGYOFDMEISGFEKGDAGVYINLSKDTFIKPVFKKIEEKKEENKPTDVSKKON 2051
QY 721 POWNHSHQNESHRKEDLQREHSHQSDSTKDYATVLDKNISSKSTTNPNK 773
DB 2052 POWNHSHQNESHRKEDLQREHSHQSDSTKDYATVLDKNISSKSTTNPNK 2104

```

RESULT 7
ADR94534 standard; protein; 637 AA.

XX ADR94534;

DT 16-DEC-2004 (first entry)

DE Novel S. pneumoniae protein sequence, SEQ ID 3169.

XX Meningitis; bacteraemia; pneumonia; otitis media; vaccine;

KM bacterial infection.

XX Streptococcus pneumoniae.

XX US6800744-B1.

XX 05-OCT-2004.

XX 30-JUN-1998; 98US-00107433.

XX 02-JUL-1997; 97US-0051553P.

XX 12-MAY-1998; 98US-0085131P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX MPI: 2004-697205/68.

XX N-PSDB; ADR91931.

PT New isolated nucleic acid encoding a Streptococcus pneumoniae polypeptide, useful for diagnosing, preventing and/or treating pathological conditions resulting from the bacterial infection.

PS Disclosure; SEQ ID NO 3169; 151pp; English.

The invention relates to an isolated nucleic acid comprising a sequence encoding a Streptococcus pneumoniae ADR91366polypeptide, or its fragments, with any of 9 fully defined sequences (appearing as ADR94308, ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682, ADR96079) or any of the fully defined sequences appearing as ADR91705, ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide sequences, or at least 40, 60 or 300 consecutive nucleotides, which is hybridizable under high stringency conditions to the nucleotide sequence. The nucleic acids and proteins are chosen from 5206 disclosed sequences. Also included are a recombinant expression vector comprising the isolated nucleic acid cited above operably linked to a transcription regulatory element, a cell comprising the recombinant expression vector and a probe comprising at least 20 consecutive nucleotides of the nucleotide sequences as cited above. The methods and compositions of the present invention are useful for the diagnosis, prevention and/or treatment of pathological conditions resulting from bacterial infection by Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and otitis media. The present sequence is one of the 2603 disclosed S. pneumoniae protein sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPRO at segdata.uspto.gov/sequence.html?docid=6800744B1.

XX Sequence 637 AA;

Query Match 75.7%; Score 3048; DB 8; Length 637;
Best Local Similarity 99.3%; Pred. No. 2.8e-168;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

QY 185 LAFAGDMELFVKDNDOKTAETIKIRMPKIKETSEYVYSSYGVNIELEGEDLSKNKPDN 244
DB 15 LAFAGDMELFVKDNDOKTAETIKIRMPKIKETSEYVYSSYGVNIELEGEDLSKNKPDN 74
QY 245 LTKWESGKIYSDSEKQYLLKDNILIRKGVALKVTYTPGKTDMLENGVYSKEDIATIQ 304
DB 75 LTKWESGKIYSDSEKQYLLKDNILIRKGVALKVTYTPGKTDMLENGVYSKEDIATIQ 134
QY 305 KANPULRLALSETTYIADSRNVEDGSTOSVMSALDGFNIIRYOVFTFRKMDKGBAIDKD 364
DB 135 KANPULRLALSETTYIADSRNVEDGSTOSVMSALDGFNIIRYOVFTFRKMDKGBAIDKD 194
QY 365 GNLVYDSSKLVLFQGDKEKYEKEDKFNVEAIKEDGSMLEFDTKPVNLSMDKNYFNPSSKN 424
DB 195 GNLVYDSSKLVLFQGDKEKYEKEDKFNVEAIKEDGSMLEFDTKPVNLSMDKNYFNPSSKN 254
QY 425 KIYVNNPEFYLRGKISDVGKGFVMELRVNESVVDNYLIIGDLHIDTRFNIKLVNKDGD 484

```


CC can be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridisation or
 CC amplification methods) also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose

XX SQ Sequence 117 AA;

Query Match 15.3%; Score 615; DB 2; Length 117;
 Best Local Similarity 100.0%; Pred. No. 3.2e-28;
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQPDGWEISGFEKGDAGYVNLKDTFIFKPKIEKKEENKPTFDVSK 716
 DB 1 YKGELEKGYQPDGWEISGFEKGDAGYVNLKDTFIFKPKIEKKEENKPTFDVSK 60

QY 717 KKDNPQVNHSQLNESHKREDLQREHRSQKSDSTKVATVLDKNNISSKSTNNPNK 773
 DB 61 KKDNPQVNHSQLNESHKREDLQREHRSQKSDSTKVATVLDKNNISSKSTNNPNK 117

RESULT 10

ABP54590
 ID ABP54590 standard; protein; 117 AA.

XX AC ABP54590;

DT 04-SEP-2002 (first entry)

XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;

XX antibacterial; Streptococcal infection; detection.

XX Streptococcus pneumoniae.

XX US2002061545-A1.

XX 23-MAY-2002.

XX 22-JAN-2001; 2001US-00765272.

XX 30-OCT-1997; 97US-00961083.

XX (CHOI/) CHOI G H.

XX (KUNS/) KUNSCH C A.

XX (BARA/) BARASH S C.

XX (DILL/) DILLON P J.

XX (DOUG/) DOUGHERTY B.

XX (FANN/) FANNON M R.

XX (ROSE/) ROSEN C A.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

XX Rosen CA;

XX WPI; 2002-479261/51.

XX N-PSDB; ABQ84825.

XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus

XX PT and for preventing or attenuating disease caused by Streptococcus

XX infection.

CC pneumoniae antigens have antibacterial activity and can be used in
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or
 CC attenuate a Streptococcal infection in an animal. The polynucleotides
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
 CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example
 CC from the present invention

XX SQ Sequence 117 AA;

Query Match 15.3%; Score 615; DB 5; Length 117;
 Best Local Similarity 100.0%; Pred. No. 3.2e-28;
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQPDGWEISGFEKGDAGYVNLKDTFIFKPKIEKKEENKPTFDVSK 716
 DB 1 YKGELEKGYQPDGWEISGFEKGDAGYVNLKDTFIFKPKIEKKEENKPTFDVSK 60

QY 717 KKDNPQVNHSQLNESHKREDLQREHRSQKSDSTKVATVLDKNNISSKSTNNPNK 773
 DB 61 KKDNPQVNHSQLNESHKREDLQREHRSQKSDSTKVATVLDKNNISSKSTNNPNK 117

RESULT 11

ADC45149
 ID ADC45149 standard; protein; 117 AA.

XX AC ADC45149;

DT 18-DEC-2003 (first entry)

XX S. pneumoniae antigenic protein SP043.

XX Antigen; bacterial infection; vaccine; pneumonia; antibacterial.

XX Streptococcus pneumoniae.

XX US6573082-B1.

XX 03-JUN-2003.

XX 28-MAR-2000; 2000US-00536784.

XX 31-OCT-1996; 96US-0029960P.

XX 30-OCT-1997; 97US-00961083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

XX Rosen CA;

XX WPI; 2003-764574/72.

XX N-PSDB; ADC45148.

XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides

XX PT useful for producing vaccines for prevention or attenuation of infection

XX by Streptococcus pneumoniae.

XX Example 1; SEQ ID NO 68; 58pp; English.

XX The invention relates to an isolated polynucleotide consisting of a

XX Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding

XX SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae

XX antigens. Also included are making a recombinant vector by inserting the

XX nucleic acid into a vector, an isolated polynucleotide consisting of at

XX least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a

XX recombinant host cell comprising the SP028 polynucleotide. The nucleic

CC acids are useful as DNA vaccine against Streptococcus pneumoniae

CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae

CC antigen nucleic acids are useful as probes for use in diagnostic methods

CC for detecting S. pneumoniae gene expression. The present sequence

CC represents an S. pneumoniae antigenic protein.

Sequence 117 AA;
Query Match 15.3%; Score 615; DB 7; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.2e-28;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQFDGWEISGFEKKDAGVYVNLSTKPTFKIKVFKKIEKKEENKPTDVS 716
DB 1 YKGELEKGYQFDGWEISGFEKKDAGVYVNLSTKPTFKIKVFKKIEKKEENKPTDVS 60
QY 717 KQDNQVHNSQVNSHREKEDLQREHSHQSDSTKDTATVLDKNTSSSTTNNPK 773
DB 61 KQDNQVHNSQVNSHREKEDLQREHSHQSDSTKDTATVLDKNTSSSTTNNPK 117

RESULT 12
ABU24813
ID ABU24813 standard; protein; 1196 AA.
AC ABU24813;
XX
XX 19-JUN-2003 (first entry)
XX
XX Protein encoded by Prokaryotic essential gene #10340.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Clostridium botulinum.
XX
XX WO20027183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX
XX (ELIT-1) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW,
XX
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreysch RA, Xu HH,
XX
XX WPI; 2003-029926/02.
XX
XX N-PSDB; ACA28683.
XX
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX
XX PT for homologous nucleic acids required for cellular proliferation to
XX
XX PT isolate candidate molecules for rational drug discovery programs.
XX
XX
XX Claim 25; SEQ ID NO 52737; 1766bp; English.

The invention relates to an isolated nucleic acid comprising any one of
the 623 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
(1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated
polypeptide or its fragment whose expression is inhibited by the
antisense nucleic acid; (4) an antibody capable of specifically binding
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
proliferation; (7) identifying a gene in an operon required for
proliferation; (8) identifying a compound that influences the activity of
the gene product or that has an activity against a biological pathway
required for proliferation, or that inhibits cellular proliferation; (9)
identifying a gene required for cellular proliferation or the biological
pathway in which a proliferation-regulated gene or its gene product lies
or a gene on which the test compound that inhibits proliferation of an
organism acts; (9) manufacturing an antibiotic; (10) profiling a
compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

Sequence 1196 AA;
Query Match 5.6%; Score 227; DB 6; Length 1196;
Best Local Similarity 20.2%; Pred. No. 0.00021;
Matches 185; Conservative 145; Mismatches 260; Indels 328; Gaps 46;

QY 1 KLGIAESKFKMLGNGKGG-----SLKDDTGVHGHQENES 39
DB 123 RLKDIQE-LFMDTGGKSGYSITIGGCKIEAVLSGPERRRSLLEBAAGIVKFKRKEAD 181
QY 40 KEKSFITIDRNISTIDPEN--KDLKKLKKKFEVDPF--TSETGRME----- 85
DB 182 KKLSTN--TEQNLIRKIDILNTYEBMEPELKESEKAKFLNSELKREVVMIYSIDK 239
QY 86 -EYKYRDDKNIIVADGTD-LEYETKLDIKSIYGVLSPSDGHPEILIGKISN 143
DB 240 IEKDLK-NISSNMLSGENIDMLKNEKSGYKEIISKF-----NEKLELIDK--NNSR 288
QY 144 NAKVYGNVYKSIIEIKATFYDFHSHKTMFEDLVANINDIVDLGAFAGDMRLPFKDDQCKA 203
DB 289 NKEBYTNKDKKNDIE-----NENALKEKIKLND-----NIKVENLKTN 331
QY 204 EIKIR-----MEKIKETSEYPPV----- 223
DB 332 EERLQKYLKEKALEKINKLKEESYLAEEIGKKEDNANNPNKELKEEKILKISBE 391
QY 224 -----SSYGVNIELGEDL--SNKKPDNLTWESGKIYS-----DSEKQOYLAK 265
DB 392 IEILNTSNTLRNEISVMEIEIVNLKLDINRSCDSYISININIKTKEDIEKEIKNIK 451
QY 266 DNIIIL-----KGYALKTYTNPQKTDMLBNGVYSKED-----IAKIOK----AN 307
DB 452 ENILLLENNKENSRISSLSKISLNKKEK-KLKEKNAAVSRLEANTHMLSNLEKHYEGYN 510
QY 308 PNIRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFRMDKGEAIDKGNL 367
DB 511 RSVKTLMEHVSYGKVDNIKGG-----CEVLGDIIVKXKELETAMEIALGAI---SNV 560
QY 368 VTDSKLVLFQKDXEYTGEDKFNVAIKEDSGMLFIDTKPVNLSMDKNYFNP--SKSN 424
DB 561 IREDE-----NKAKILINYLKCK-----SLGRATFPPLTTIOQR 594
QY 425 KTYVNPPEFYLAGKISDKGFWMLRVNSESVDNVLIVGDLHIIDTRDNI--KLVNKG 482
DB 595 KAKINN-----VIREDF--GLIASDLIDY-----DVFSNIDIVLGRITLAKO- 637
QY 483 DLMGDMKQKANGFPDKYTDMDGNVYVLTQGYSDLNKAVGVHYOFLYDNVPEVNDPK 542
DB 638 --MDSALTKAKLANSFKIVLIEGV-----ING 665
QY 543 GNT---SIEYADGKSVFNINDKRNNGPDGEIOEHYIYNGKEYSPNDIKQIIDKTLNI 599
DB 666 GSLTGSGISGRAGSSI--ISRR-----EIEE-----TKLELETKN- 700
QY 600 KIVVDPANVTYVKKFLLNKDNGEVSSELKHHVVTYTIQNGKEMSSYVSEDFIIPVYG 659
DB 701 -----TIEEWGN-----ILEKNKIKY--LDEENLNI---KD 728
QY 660 ELEKGYQFDGWEISGFEKKDA-----GVINLSKD--TFIKVPFKKIEKKEENK 709

CC the native pattern of folding. Larger amounts of the protein can be
CC produced recombinantly than would be possible using the parasites as
CC source. (Updated on 25-MAR-2003 to correct PR field.)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:10:40 ; Search time 45 Seconds
(without alignments)
1282.305 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGEIAESKFKVLGNKGESG.....ATVLDKNNISSKSTNNPNK 773

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues.

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents, AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/ECTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3789.5	94.1	2138	4 US-09-583-110-5274	Sequence 5274, Ap
2	3048	75.7	637	4 US-09-107-433-3169	Sequence 3169, Ap
3	750.5	18.6	1529	4 US-09-107-433-4771	Sequence 4771, Ap
4	615	15.3	117	3 US-08-961-083-68	Sequence 68, Appl
5	615	15.3	117	4 US-09-536-784-68	Sequence 68, Appl
6	203	5.0	1166	4 US-09-200-650E-7	Sequence 7, Appl
7	200.5	5.0	1786	3 US-08-973-462-8	Sequence 8, Appl
8	191.5	4.8	2504	4 US-09-328-352-5821	Sequence 5821, Ap
9	188.5	4.7	670	4 US-09-107-433-4976	Sequence 4976, Ap
10	188.5	4.7	708	4 US-09-583-110-3019	Sequence 3019, Ap
11	187.5	4.7	1529	2 US-08-728-470-10	Sequence 10, Appl
12	187.5	4.7	1529	3 US-08-719-641-10	Sequence 10, Appl
13	186.5	4.6	930	4 US-09-200-650E-3	Sequence 3, Appl
14	186	4.6	1315	4 US-09-800-650E-5	Sequence 5, Appl
15	185	4.6	2733	4 US-09-949-016-11433	Sequence 11433, A
16	185	4.6	3359	4 US-09-949-016-6507	Sequence 6507, A
17	184	4.6	1588	5 PCT-US93-07261-11	Sequence 11, Appl
18	184	4.6	1663	5 PCT-US93-07261-16	Sequence 16, Appl
19	181.5	4.5	778	3 US-09-134-001C-3868	Sequence 3868, Ap
20	179.5	4.5	746	4 US-09-710-279-652	Sequence 652, Appl
21	179.5	4.5	1183	2 US-08-447-031A-2	Sequence 2, Appl
22	179	4.4	2142	4 US-09-540-236-3459	Sequence 3459, Ap
23	178	4.4	2375	4 US-09-538-092-1131	Sequence 1131, Ap
24	177.5	4.4	1600	2 US-08-617-697-10	Sequence 10, Appl
25	177	4.4	800	6 5183745-3	Patent No. 5183745
26	177	4.4	800	6 5183745-3	Patent No. 5183745
27	174.5	4.3	1073	4 US-09-206-942-49	Sequence 49, Appl

28	174.5	4.3	1079	4 US-09-206-942-47	Sequence 47, Appl
29	174.5	4.3	10182	3 US-09-134-001C-3159	Sequence 3159, Ap
30	174	4.3	3070	4 US-09-961-403-7	Sequence 7, Appl
31	173.5	4.3	2253	3 US-09-377-155-33	Sequence 33, Appl
32	173.5	4.3	2353	3 US-08-913-942-4	Sequence 4, Appl
33	173.5	4.3	2353	3 US-09-669-974-33	Sequence 33, Appl
34	173.5	4.3	2353	4 US-09-797-862-33	Sequence 33, Appl
35	173.5	4.3	2353	4 US-09-684-707-4	Sequence 4, Appl
36	173.5	4.3	2411	3 US-09-268-347-36	Sequence 36, Appl
37	173	4.3	930	3 US-09-134-001C-5314	Sequence 5314, Ap
38	173	4.3	930	4 US-09-386-962C-10	Sequence 10, Appl
39	173	4.3	2710	1 US-08-480-604A-6	Sequence 6, Appl
40	173	4.3	2710	2 US-08-405-496A-6	Sequence 6, Appl
41	173	4.3	2710	3 US-08-915-136-6	Sequence 6, Appl
42	173	4.3	2710	3 US-08-957-310-6	Sequence 6, Appl
43	173	4.3	2710	4 US-10-011-366-6	Sequence 6, Appl
44	173	4.3	2710	4 US-09-084-517-6	Sequence 6, Appl
45	172.5	4.3	1092	4 US-09-147-405B-15	Sequence 15, Appl

ALIGNMENTS

```
RESULT 1
US-09-583-110-5274
Sequence 5274, Application US/09583110
Patent No. 6699703
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
FILE REFERENCE: PAT00-07A
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US/09/583,110
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 5274
LENGTH: 2138
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-583-110-5274

Query Match          94.1%; Score 3789.5; DB 4; Length 2138;
Best Local Similarity 94.8%; Pred. No. 1.1e-238;
Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;

1 KLGEIAESKFKVLGNKGESGSLKKDTTGVYHHQENEBSEKSSFTIDRNTTIDFENK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1333 KLGEIPESKFKVLGNKVDLSLNKEFAEVENMLVDNOSIBGKSLFNHKTSTIDFENK 1392
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 DLKKLILKKKFRVDFDFTSETKRMEYDYKYVDKGNIIAYVDGTDLEAYETKLDSEIKSI 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1399 DLKKLILKKKFKKEDFPVNGGRTTYERDYKDDKGNIIAYVDGTDLEAYETKLDSEIKSI 1451
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 YGVLSPSKDGAFELIIGKISNVSKNAKYVYNNYKSIIEIKATYDFHSHKTMFDLYANIND 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1452 YGVLSPSKDGAFELIIGKISNVSKNAKYVYNNYKSIIEIKATYDFHSHKTMFDLYANIND 1511
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 IYDGLAFADMDLFPYKNDQDKAAEIKIMPEKIKETKSEYFPVSSYGVNIEIGEDPLSN 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1512 IYDGLAFADMDLFPYKNDQDKAAEIKIMPEKIKETKSEYFPVSSYGVNIEIGEDPLSN 1571
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 KPDNLTIKMESGKIYSDSEKQYLLKNDIILKRGVALKTTYVPGKTDMLEGNGVYSKEDI 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1572 KPDNLTIKMESGKIYSDSEKQYLLKNDIILKRGVALKTTYVPGKTDMLEGNGVYSKEDI 1631
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 AKIQKAPNPLRALSETTYIADSRNVEDGRSTQSVLMSALDGFNIIRYQVFFKNDKGBA 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```


Db 1632 AKIQANPVLRLSETTIYADSRNVEDGRSTQAVLMSALDGNIRIYQVFTFKANDKGEA 1691
Qy 361 IDKDNLTVDSSKLVLFCKDKEKYTGEDKFNVEA1KEDGSMFLIDTKPVNLSMDKNYFNP 420
Db 1692 IDKDNLTVDSSKLVLFCKDKEKYTGEDKFNVEA1KEDGSMFLIDTKPVNLSMDKNYFNP 1751
Qy 421 SKSNKTIYRNPEFYLRGKISDGKGFNWEIRVNESVVDNLIYGDHINDTRPFIKLVNK 480
Db 1752 SKSNKTIYRNPEFYLRGKISDGKGFNWEIRVNESVVDNLIYGDHINDTRPFIKLVNK 1811
Qy 481 DGDINDMGKDYKANGFPDVTMDGNVYLQTSYGLNKAAGVHYQFLYDNVKEPVNID 540
Db 1812 DGDINDMGKDYKANGFPDVTMDGNVYLQTSYGLNKAAGVHYQFLYDNVKEPVNID 1871
Qy 541 PKGNTSIEYADGKSVFVNINDKRNNGPDGEIOEHYIYNGKEYTSFNDIKQIIDKTLN1K 600
Db 1872 PKGNTSIEYADGKSVFVNINDKRNNGPDGEIOEHYIYNGKEYTSFNDIKQIIDKTLN1K 1931
Qy 601 IIVKDPARTTYKEPILNKDNGEVSEIKPHRTVTYTIQNGKEMSTTVSEDFILPVYKGE 660
Db 1932 IIVKDPARTTYKEPILNKDNGEVSEIKPHRTVTYTIQNGKEMSTTVSEDFILPVYKGE 1991
Qy 661 LEKGYPDGMWISGEGKKDAGVYNLSKDTFIKPVFKKIEKKEEENKPTDVSKKDN 720
Db 1992 LEKGYPDGMWISGEGKKDAGVYNLSKDTFIKPVFKKIEKKEEENKPTDVSKKDN 2051
Qy 721 PQVNSQLNESHKEDLQREHSHQSKSDSTKDVATATVLDKNNISKSXTNNPNK 773
Db 2052 PQVNSQLNESHKEDLQREHSHQSKSDSTKDVATATVLDKNNISKSXTNNPNK 2104

RESULT 2
US-09-107-433-3169
; Sequence 3169, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID and AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: Protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...637
SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-09-107-433-3169
Query Match 75.7%; Score 3048; DA 4; Length 637;
Best Local Similarity 99.3%; Pred. No. 5,6e-191;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Qy 185 LAFAGDMFLFYKNDQKAAEIKIRPEKIKETKSEHYPVSSYGVNIELEGDLSPNKPDN 244
Db 15 LAFAGDMFLFYKNDQKAAEIKIRPEKIKETKSEHYPVSSYGVNIELEGDLSPNKPDN 74
Qy 245 LTKWESGKIYDSEKQOYLNDNIIIRGYALAKTTYNPKTMDLENGVYSKEDIATIQ 304
Db 75 LTKWESGKIYDSEKQOYLNDNIIIRGYALAKTTYNPKTMDLENGVYSKEDIATIQ 134
Qy 305 KANPULRALSETTIYADSRNVEDGRSTQAVLMSALDGNIRIYQVFTFKANDKGEA1DKD 364
Db 135 KANPULRALSETTIYADSRNVEDGRSTQAVLMSALDGNIRIYQVFTFKANDKGEA1DKD 194
Qy 365 GNLVTDSSKLVLFCKDKEKYTGEDKFNVEA1KEDGSMFLIDTKPVNLSMDKNYFNP 424
Db 195 GNLVTDSSKLVLFCKDKEKYTGEDKFNVEA1KEDGSMFLIDTKPVNLSMDKNYFNP 254
Qy 425 KIYRNPEFYLRGKISDGKGFNWEIRVNESVVDNLIYGDHINDTRPFIKLVNKDGI 484
Db 255 KIYRNPEFYLRGKISDGKGFNWEIRVNESVVDNLIYGDHINDTRPFIKLVNKDGI 314
Qy 485 MDGMKDYKANGFPDVTMDGNVYLQTSYGLNKAAGVHYQFLYDNVKEPVNIDPKGN 544
Db 315 MDGMKDYKANGFPDVTMDGNVYLQTSYGLNKAAGVHYQFLYDNVKEPVNIDPKGN 374
Qy 545 TSIEYADGKSVFVNINDKRNNGPDGEIOEHYIYNGKEYTSFNDIKQIIDKTLN1K 604
Db 375 TSIEYADGKSVFVNINDKRNNGPDGEIOEHYIYNGKEYTSFNDIKQIIDKTLN1K 434
Qy 605 DPAVTTYKEPILNKDNGEVSEIKPHRTVTYTIQNGKEMSTTVSEDFILPVYKELKG 664
Db 435 DPAVTTYKEPILNKDNGEVSEIKPHRTVTYTIQNGKEMSTTVSEDFILPVYKELKG 494
Qy 665 YQFDGMEISGEGKKDAGVYNLSKDTFIKPVFKKIEKKEEENKPTDVSKKDNPOVN 724
Db 495 YQFDGMEISGEGKKDAGVYNLSKDTFIKPVFKKIEKKEEENKPTDVSKKDNPOVN 554
Qy 725 HSQLNESHKEDLQREHSHQSKSDSTKDVATATVLDKNNISKSXTNNPNK 773
Db 555 HSQLNESHKEDLQREHSHQSKSDSTKDVATATVLDKNNISKSXTNNPNK 603

RESULT 3
US-09-107-433-4771
; Sequence 4771, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID and AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:


```

      MEDIUM TYPE: CD/ROM ISO9660
      COMPUTER: <Unknown>
      OPERATING SYSTEM: <Unknown>
      SOFTWARE: <Unknown>
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/09/107,433
      FILING DATE: 30-Jun-1998
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 60/085131
      FILING DATE: May 12, 1998
      APPLICATION NUMBER: 60/051553
      FILING DATE: July 2, 1997
      ATTORNEY/AGENT INFORMATION:
      NAME: Ariniello, Pamela Deneke
      REGISTRATION NUMBER: 40,489
      REFERENCE/DOCKET NUMBER: GTC-011
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (781)893-5007
      TELEFAX: (781)893-8277
      INFORMATION FOR SEQ ID NO: 4771:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 1529 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      HYPOTHEetical: YES
      ORIGINAL SOURCE:
      ORGANISM: Streptococcus pneumoniae
      FEATURE:
      NAME/KEY: misc feature
      LOCATION: (B) LOCATION 1...1529
      SEQUENCE DESCRIPTION: SEQ ID NO: 4771:
US-09-107-433-4771
      Query Match      18.6%; Score 750.5; DB 4; Length 1529;
      Best Local Similarity 79.7%; Pred. No. 1.8e-40;
      Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

QY      1 KGEIAESKFNKNGSGSLKKDTGTVHHHQNESIKESFTIDNINISIRDFENK 60
      |||||
DB      1336 KGEIPESKFNKNGSKDLSLNKETAEVNNLLVDQSLIEGSLFNIHKITISIRDFENK 1395
      |||||
QY      61 DKKLIKRRREVDFSETGRKMEEDYKYDDKGNIIAYDGTDLVEYTEKLDIEIKSKI 120
      |||||
DB      1396 DKKLIKRRREVDFSETGRKMEEDYKYDDKGNIIAYDGTDLVEYTEKLDIEIKSKI 1454
      |||||
QY      121 YGVLSKSGKGFPEILKISNVSKNAVYVGNNTKSLIEIKATKDFHSKMTPLVYANIND 180
      |||||
DB      1455 YGVLSKSGKGFPEILKISNVSKNAVYVGNNTKSLIEIKATKDFHSKMTPLVYANIND 1514
      |||||
QY      181 IYDGLAF 187
      |||||
DB      1515 IYDGLAF 1521
      |||||

RESULT 4
US-08-961-083-68
      Sequence 68, Application US/08961083
      Patent No. 6159469
      GENERAL INFORMATION:
      APPLICANT: Choi et. al.
      TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
      NUMBER OF SEQUENCES: 452
      CORRESPONDENCE ADDRESS:
      ADDRESSER: Human Genome Sciences, Inc.
      STREET: 9410 Key West Avenue
      CITY: Rockville
      STATE: Maryland
      COUNTRY: USA
      ZIP: 20850
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
      COMPUTER: HP Vectra 486/33
```

```

      OPERATING SYSTEM: MSDOS version 6.2
      SOFTWARE: ASCII Text
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/961,083
      FILING DATE:
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER:
      FILING DATE:
      ATTORNEY/AGENT INFORMATION:
      NAME: Brookes, A. Andreb
      REGISTRATION NUMBER: 36,373
      REFERENCE/DOCKET NUMBER: PB340P2
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (301) 309-8504
      TELEFAX: (301) 309-8512
      INFORMATION FOR SEQ ID NO: 68:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 117 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: protein
US-08-961-083-68
      Query Match      15.3%; Score 615; DB 3; Length 117;
      Best Local Similarity 100.0%; Pred. No. 3.9e-33;
      Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      657 YGGELEKGYQFDGWEISGFEGRKDAVYINISKDTFIKVPFKIEKKEENKPTFDVSK 716
      |||||
DB      1 YGGELEKGYQFDGWEISGFEGRKDAVYINISKDTFIKVPFKIEKKEENKPTFDVSK 60
      |||||
QY      717 KQDNPVNSQVNSHREKEDLOREHSQKSDSKDTQVTAIVLDKNNISSKSTNNPK 773
      |||||
DB      61 KQDNPVNSQVNSHREKEDLOREHSQKSDSKDTQVTAIVLDKNNISSKSTNNPK 117
      |||||

RESULT 5
US-09-536-784-68
      Sequence 68, Application US/09536784
      Patent No. 6573082
      GENERAL INFORMATION:
      APPLICANT: Choi et. al.
      TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
      NUMBER OF SEQUENCES: 452
      CORRESPONDENCE ADDRESS:
      ADDRESSER: Human Genome Sciences, Inc.
      STREET: 9410 Key West Avenue
      CITY: Rockville
      STATE: Maryland
      COUNTRY: USA
      ZIP: 20850
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
      COMPUTER: HP Vectra 486/33
      OPERATING SYSTEM: MSDOS version 6.2
      SOFTWARE: ASCII Text
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/09/536,784
      FILING DATE: 30-Oct-1997
      CLASSIFICATION: <Unknown>
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 08/961,083
      FILING DATE: OCT-30-1997
      ATTORNEY/AGENT INFORMATION:
      NAME: Michelle S. Marks
      REGISTRATION NUMBER: 41,971
      REFERENCE/DOCKET NUMBER: PB340P3
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (301) 309-8504
      TELEFAX: (301) 309-8512
      INFORMATION FOR SEQ ID NO: 68:
```

```

;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 117 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match      15.3%; Score 615; DB 4; Length 117;
Best Local Similarity 100.0%; Pred. No. 3,9e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQPDGWEISGFEKGDAGYVNLSTKDTPIKVFKEKKEEENKPTFDVSK 716
DB 1 YKGELEKGYQPDGWEISGFEKGDAGYVNLSTKDTPIKVFKEKKEEENKPTFDVSK 60
QY 717 KKDNPVNHSQLNEHSHKEDLQREHSHSQSDSTKDTATVLDKNNISSKSTNNPNK 773
DB 61 KKDNPVNHSQLNEHSHKEDLQREHSHSQSDSTKDTATVLDKNNISSKSTNNPNK 117

RESULT 6
US-09-200-650E-7
; Sequence 7, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
;   APPLICANT: Patti, Joseph M.
;   APPLICANT: Foster, Timothy J.
;   APPLICANT: Hook, Magnus A.O.
;   APPLICANT: Eidiim, Deidre M.
;   APPLICANT: Perkins, Samuel L.
;   TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
;   FILE REFERENCE: P06283US2/BAS
;   CURRENT APPLICATION NUMBER: US/09/200,650E
;   PRIOR FILING DATE: 1998-11-25
;   PRIOR APPLICATION NUMBER: 60/066,815
;   PRIOR FILING DATE: 1997-11-26
;   PRIOR APPLICATION NUMBER: 60/098,427
;   PRIOR FILING DATE: 1998-08-31
;   NUMBER OF SEQ ID NOS: 23
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO: 7
;   LENGTH: 1166
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-200-650E-7

Query Match      5.0%; Score 203; DB 4; Length 1166;
Best Local Similarity 21.6%; Pred. No. 6,8e-05;
Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

QY 33 QENESIKE--KSFIDRNISTIRDFEKDKLKKKKFR-----EYDDFTS 78
DB 226 KNPPEKLEIVNDSNTDSTPAPATAPSAFKVNAKMRRAVAVQAPPAVANNNDLIX 285
QY 79 ETGKMEEYDYKDKNIIAVDGTDLLEYETE-----KLDEIKSKIYGVLS 125
DB 286 VTQKQIK---VGDKDNVAAAHADGDIETDFTIDNKVKKGDITMIVYDKNVI.PSDLT 341
QY 126 PSKD-----GHFEILGKISNVGNKAK---VYGNMYSKLEIKATKYDHSK----- 168
DB 342 DKNDIDITDTPSGEVIKAGTFPKATKQITTYFTDVKDKEDIKSRLLTYSYIDKKTVPNE 401
QY 169 ---TWTF-----DLVANNI-DIVDGLAFAGD---RELFEVK-DNDQKAEIKIRM-PEKIX 214
DB 402 TSLNITFATAGKETSQNVTVYQDPMVH-GDSNISOIFKLDEDEKQTIHQIYVNPPLKKS 460
QY 215 ETKSEY---PYVSSYGVNIEIGSD--LSKNKPPNLTQM-----ESGKIYDSEKQ 261
DB 461 ATNTKYDIAGSQVDVYGN-IKLNGSTIIDQNTETIKVYKVNSDQQLPSQNR.IYDPSQYED 519
QY 262 YLLK-DNIIIRKGYALKTTTYPGKTIDMEGNGVSKEDIKIQANPMLALSTETIYA 320
```

```

DB 520 VTSQPDN---KKSFSNNVATLDFG-----DINSAT-----IKVSKYTPTS 558
QY 321 DSR-NVEDGRSTQSVLMSALDGFNIIRQVTFPKANDKG-----E 359
DB 559 DGEELDIAGTSMRTT--DKGYVYVAGSYNSITVNSDNGGGDGYVKEPEKLYKIGDYWE 616
QY 360 AIDKDNLVTSKVLFGKDKDEYTGDKKNVEAIKEDGSMFLFDIYKPNLSMDKNY-F 418
DB 617 DVDKDG-----VGDTDSKEKPMANVL-VLTLYPDGT-----TKSVRTDANGHYEF 660
QY 419 NPSKSNKIYV---RNPEFYLNKGIS-----DKGFNWEILRVNESVVDNLIYGDLLHD 468
DB 661 GGLKDGERTYVKEFTPTGYLPTKVNQTTDGEKDSNGSSVYKINK-----DDMSLD 712
QY 469 -----NTRDENIKILNVKG--DIMDMQKDYKANGPDKYVDMGNVYLQTVGSDL 517
DB 713 TGFYKPKRYNLDGYWEDTNDGIDQANPEIKQVYV-----LKDSIGKV-IGTTTIDA 766
QY 518 NAKAVGVHYQFLYDNVKEVNIIDPKNTSIEY---ADGKSVFN--INDKRNPFDSIQ 572
DB 767 SGK-----YKF-----TDLD-NGNYTVEFETPAGYTPYKQTTADDKDSNGL----- 807
QY 573 EGHYIYNGKEYTSFNDIKQIIDKTLNLIKIVKDPARN--TYKRI---LAKDGEVSEL 627
DB 808 -----TTGVADADNMTLD---RGFYKTPKYSGLDGVWYDSNKGKQDSTE 851
QY 628 KPHR-VTVTIONGK-EMSSITVSEEDFILPYKGELEKGYQPDGWEISGFEKGDAGYVI 685
DB 852 KGIDVTVTLNKEGEVIGTKIDEN-----GK-----YFSD----- 883
QY 666 NLSKDTIKPYFKKIEEKEEENKPTFVSKKKDNPQV-----NHSQ--LNEHSHKEDLQ 738
DB 884 NLDSEKY-KVIFPKAGLQVTVTTTBD-DKDADGGEVDVITDHDFTLNGYFEEDTS 941
QY 739 REHSHQKSDSTKD 751
DB 942 DSDSDSDSDSDSD 954

RESULT 7
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;   APPLICANT: DRUILLHE, PIERRE
;   APPLICANT: DAUBERSIES, PIERRE
;   TITLE OF INVENTION: MALARIAL PRE-BRYTHOCYTIC STAGE POLYPEPTIDE MOLECULES
;   FILE REFERENCE: 0660-0125-0 PCT
;   CURRENT FILING DATE: 1998-02-06
;   EARLIER APPLICATION NUMBER: PCT/FR96/00894
;   EARLIER FILING DATE: 1996-06-12
;   EARLIER APPLICATION NUMBER: FR 95/07007
;   NUMBER OF SEQ ID NOS: 29
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO: 8
;   LENGTH: 1786
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match      5.0%; Score 200.5; DB 3; Length 1786;
Best Local Similarity 20.5%; Pred. No. 0.00018;
Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

QY 6 AES--KFRN-LANGKESGLKKDT-----TGVEHHHQENE-----ESIK 40
DB 868 AESVTTFNIIIEIQENTITNTDIEKLELHENVLSALNTQSEKKEKVIDVIEEVK 927
```



```

0Y 679 KDAGV---ILNSDFTFK-PVFKKIEKKEENKPFDPVK-----KCDPQV 723
Db 2177 KQTNVADGN-ADAVNKGQLDKQINEVKQDQIGK---DIGKSDHVAVDKDKNGVNDK 2232
0Y 724 NNSQLNESHREDDQREHESQKSDSTYKDV-----ATVLDKNNISKSTNN 770
Db 2233 NSVTLGGGEKGTNKLKNVADKRVASGSDAVNAGGLMNIQNVDKNSNDIKNIQNN 2287

RESULT 9
US-09-107-433-4976
; Sequence 4976, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4976:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 670 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEtical: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...670
; SEQUENCE DESCRIPTION: SEQ ID NO: 4976:
US-09-107-433-4976

Query Match 4.7%; Score 188.5; DB 4; Length 670;
Best Local Similarity 21.4%; Pred. No. 0.00029;
Matches 138; Conservative 92; Mismatches 233; Indels 183; Gaps 33;

0Y 223 VSSYGVNVEIEBGLSKNKPDLTKMESGKIYDSEKQY---LKKNIILR--KGYALK 277
Db 54 ISSNGTI-----RSNSQLDNR-T-VTSITNENKSYKEDVISRIIIEKEDNALS 104
0Y 278 VTTN-----PGKTMLENGVYSKEDIKIQ-----KANPVLRLASE 315
Db 105 VKDVGAVGDGIHDDRQAIQDAIDAAOGLGGANVYFPEGTVLVEIYVFLKSHTHLEINER 164

```

```

QY 316 TTYIADSNVDDGDSSTGVSMLSDPFGNIIKRYQVTFPFKANDKGALIDDGMLVYDSSKLY 375
Db 165 ATI-----LNGINIKNHPISIVFT-----GLFTDGAQV 193
QY 376 LFG-KDDEYTGDEGFENVAIKEDGSMFLFDYKPYNLGM--DKYFNPSKNIYVRNP 432
Db 194 EMGFREDISTSGGTIDNMGNALNEB-----TKACNLILINSSGAFALGNSNVTIKVY 247
QY 433 F---YLKGISDKGGFNNELRVNESVDVNYLIYGDHLIDNTRDENIKLVKDGDIIM--D 486
Db 248 FKDSYGHAIQIAGSKN-----VLVNSRFLGA-----LPYMKDGGQIISKES 291
QY 487 WGMKDYKANGEPDKYTDMDG---NVLQOTY---SDLNAM---AVGHYGEIYDYNVKE 536
Db 292 IQIELETRKGFYPYALND-DGKSEVYTIQNSIFGKSDSGELVYALGTHYQIQLTQNPDSN 350
QY 537 VNIDPKGNTSIEYADGKSVENINDKRNNGFDEGIOEHY-----INGKEYTSFND 588
Db 351 IKILNNHDDNMVYAGVRFGTGFDVLKGNRPDKKYGSGSHYRESGALVYNAVYKNTKD 410
QY 589 I---KQII-----DKTLNKIYVKOPAR-----NTYKEFLNKDTGVSSEIK 628
Db 411 LLDLNKQVVAIENTFNINADPKTKAIR-VAKDSAEYLGKVSITVYKVNINNSKTEQ-- 467
QY 629 PHRTVVTIQNKEKMSSTIV--SEEDFLIPYKGEI---BKGYQDGEI-----SGPBG 678
Db 468 FNIELLRSDLVVSENSIFGKGEIVIEDSKGKITVLNNGFYNLGKIYISYISNANGK 527
QY 679 K-----DAGVIVLTKDTPFKPVFKKIEKKKEE-----NKPTFDVSK 716
Db 528 EPVLRDSDGNFNI-VTENGLYKIVTNLNSDKNEKKNKEKQYNSNNVYDINSQKNGEFTNS 586
QY 717 KKNPQVNVHSQNLNESHK-----EDIQREHSQKSDSTQV 752
Db 587 SKDNKQNDKIDNKQDNKTEBEVNYKIVGDGRETEHINKSKXIYDV 632

RESULT 10
US-09-583-110-3019
; Sequence 3019, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: P4TH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3019
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-3019

Query Match 4.7%; Score 188.5; DB 4; Length 708;
Best Local Similarity 21.4%; Pred. No. 0.00031;
Matches 138; Conservative 92; Mismatches 233; Indels 183; Gaps 33;

QY 223 VSSYGVNVELAGEGLSKNKNPNLTKMESGKIYSDSEKQY--LLKQNIILR--KGVALK 277
Db 92 ISSNGTI-----RNSQDLNRT-VES-TVTSINENKSYKEDVYSDIILKKEFEDTALS 142
QY 278 VTTYN-----PGTDTMLENGVYSKEDIKIQ-----KANPILPAISE 315
Db 143 VKDGVANGDGIHDDRQALQDAIDAAOGILGGANVFPFGTYLVKEIVFLKSTHLELNK 202

```

```

QY 316 TTIVADSRNVEDGRSTQSVLMSALDGFNIIRYOVFTKRNKDGEALDKDGNLVTDSKLV 375
DB 203 AII-----LNCINIKNHSIVFTM-----GLFTDGAQV 231
QY 376 LFG-KDKEKYTEDEKFNVEAIKEDGSMLEPDKPVNLSN--DKNVPNSKNIYVRNE 432
DB 232 EWGPEFEDISYSGGTIDMNGALNBEG-----TKAKVLPLINSQALGAINSNVITIKVNT 285
QY 433 F---YLKGISDKGFGNFMELRVNESVVDVLLYGDHLHINTDPMIKLVKOGDIN--D 486
DB 286 FDSYQGHAIQIAGSKN-----VLYDNRFLGQA-----LPKTKGQGLIISKES 329
QY 487 WGMKDYKANGPPDKVYTDMDG-----NYLQDTGY---SDLNAK-----AVGVHYQFLYDVKPE 536
DB 330 IQIEPLTRKGFYALAND--DKKSENVATIONSYFGSKDSKSGELVTAIGTHQTLSTONPSN 388
QY 537 VNIDPKGNTSIEVADKSVVFVFNINDKRNNGFDEIQEHIY-----INGKEYTSFND 588
DB 389 IKILNNHFDNMVYAGVRFGTVDVLKGNRFDKVKVGSVHRESGALVNVASYKNTKD 448
QY 589 I-----KQIT-----DKTANIKIYKDPAR-----NTYKSEFILAKDTGEVSEIK 628
DB 449 LLDLNRKQVIAENIFNIADPKTKAIR--VAKDSAEYLKGVSDIYVTKVNNINSKETEQ-- 505
QY 629 PRRVVTIIONGKEMSSTIV--SEEDFILPVYKGEI-----EKGYQFDGWEI-----SGFEKG 678
DB 506 FVIELLRVNDNLVVSNSIFPGKEGIVIEDSKCKITVLANQFYNLSKGISPFKSNANK 565
QY 679 K-----DAGVYINLSKDTFIKPVFKIEKKEE-----NKPTDVSAR 716
DB 566 EPIYRSDGNFNI--VTENGLYKLVTNLSDKNKEKKEKKEKQYNSNVNIDSNQKGEFNS 624
QY 717 KDNQPVNHSQLNESHRK-----EDLQREESQKSDSTKQV 752
DB 625 SKDNQMDNDKIDNKQDNKTEEVYKIVGGRTEHNIINSKELVDV 670

RESULT 11
US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: OF No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Berktesseer, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0810
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-728-470-10

Query Match 4.7%; Score 187.5; DB 2; Length 1529;
Best Local Similarity 20.1%; Pred. No. 0.001;
Matches 148; Conservative 94; Mismatches 255; Indels 239; Gaps 37;

QY 9 KFKNL---GNGKEGSLKQDTGVENHHOENESIKESFTIDRNISTRIDFENKDLK-- 63
DB 531 RKNVNSINGTK-----GLKFIANQNNFTKFDGELNIS--GIVTINQTKDKVKKW 580
QY 64 KLIKKKFREVDFTSETGRMEEDY--KYDDKGNITAYDDGTDLLEYTEKLEIKSIYV 122
DB 581 NAKSOSYNNVSLTINT--VQKFTIKFVDSGS-----NSQDLSSRRSPAG 625
QY 123 VLSPSKQGF--ILKIS--NVSNAKAVY-----GNKYSEI---KATKYPDSHT 169
DB 626 V-----HFNGIGKTFNFIQANAKALFKLPNATDPKELPTFNANITATGSDS 678
QY 170 MTFEDLYANIND-----IYDGLAFADMDLFLVKNDOCKAEIKIRMEPIKETSEVPYVS 224
DB 679 VMFDIHANLTSAAIGIMDSITITGIDPSITSHRNSNAFEIKDITINAGSINS-- 735
QY 225 SYGVNIEIGEDGLSKNKNPNLTVMESGKIYSDSEKQYLLKONILLRKGYALKVTTYNDG 284
DB 736 -----LQTKDSFYNEYSKHALINSNHLFT----- 760
QY 285 KTDMEKNGVYSKEDIQKANKPULRALSETTIYVADSRNVEDGRSTQSVLMSALDGFNI 344
DB 761 ---LGNVTLLGESSSSITGNITNKNANTLQADTSNSNTGLKRRLLTLG----- 809
QY 345 IRYVFTFGMDNKGALIDKGNLVDSKLVFGK---DDKEYTEDEKFNVEAIKEDGS 400
DB 810 -----NISVENSLTGANANIVGNLSIAEDTFKGE----- 841
QY 401 MLEFIDTKPVNLSMDKNVFNPSKNIYVNPFEYLRGKISDKGFNEMELRVNESVVDNVL 460
DB 842 -----ASDNLNITGFTNNGTAN--INIKQGVVKLQGDINNKGGLN--ITTNASGTOKTI 892
QY 461 IYGDHLHINTDPMIKLVK--DGDIMDG-----MKYKAN-----G 496
DB 893 INGNI--TNEKGLNLIK--NIKADAEIQIGNISQKGNLTISSDKVNTIQITIKAVEGG 950
QY 497 FPDKYTMDMGVNYLOT-----GY-----SDL-----NAKAVGVHY 526
DB 951 RSDSSEAEANALTIQKELKLAGDLNISGFNAKITAKNGSDLTITGNASGNADAKV-- 1008
QY 527 QFLYDVKPEVNIIDPKGNTSIEVADKSVVFVFNINDKRNNGFD--GEIQEHIYINGKEYT 584
DB 1009 ---TFDKVK-----DSKIST-----DOHNVTLNSEVTSNGSSNAGDNGSTGLTISAKDVT 1056
QY 585 SPNDIKQIIDKTLNLIKIVKDFARNTVVE--FILKDTGEVSEIKPRVTVYONG----- 639
DB 1057 VNNNVTS--HKTINSA--AGNVTKEGTINATGTSVE-----VTAQNTIKNG 1102
QY 640 ---KEMSTIVSEEDFI 653
DB 1103 NITSQNVTVTATENLV 1118

RESULT 12

```

US-08-719-641-10
Sequence 10, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: PC floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkesreaser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 4.7%; Score 187.5; DB 3; Length 1529;
Best Local Similarity 20.1%; Pred. No. 0.001;
Matches 148; Conservative 94; Mismatches 255; Indels 239; Gaps 37;

QY 9 KEKNT---GNGKESLKKOTGVHHHNEESIKESFTIDRNSTIRDPENOLK-- 63
DB 531 RNNVNSINTGK-----GKPIANNNFTHKFDGLNIS-GVITINOTTKDVKW 580
QY 64 KLIKKKFREVDFTSETGRMBEYD-KYDDKGNIIAYDGDLEVEYEKLDIKSKIY 122
DB 581 NASKOSYNNVSLTNT---VQKFTPIKFDVDSG-----NSQDLSSRRSFAG 625
QY 123 VLSPEKDGHE-ILKGIS-NVSKNAKVY-----GNNYKSIET---KATKIDPHSKT 169
DB 636 V-----HENGIGGKTNNIGANNAALFKLPNATDKKELPTTFNANITATGNSDS 678
QY 170 MTFEDLVANIND-----IVDGLAFAGDMRLFVKNDODKKAELIKIRMEKIKETSEYFVS 224
DB 679 VWFIDHANLTSRAAGINMSINITGSLDSITSHRNSNAFEIKDOLITNATGSNFS--- 735
QY 225 SYGVNIEIGEGDLSKNNKPNLTMRSGKIYSDSEKQOYLKKNIIILRKGYALKVTTYNG 284
DB 736 -----LKQTKDSFYNEYSKKAINSNHLRT----- 760

QY 285 KTDMEGNGVYSKEDIKAKIKANPNLRALSETTIYADSRNVEDGRSTQSLMADGFNI 344
DB 761 ---LGGNVTLGEGSSSSITGNITNKRNAVTLQADTSNSNTGLKKTLLTG----- 809
QY 345 IRYQVFTFMDKGBAIDKGNLVTDSSKLVLFGR-----DDEKTYGDEKFNVEAIKEDGS 400
DB 810 -----NISEGVLSTGANANIVGSLTAEDETFGE----- 841
QY 401 MFLDTKPNVLSMDKNVFNPSKNIYVNPFEYLRGKISDGKGFNMBELRVNESVVDNYL 460
DB 842 -----ASDNLINITGTFNNGTAN-INIKQGVAKQDGINNKGILN-ITTNASGTQKTI 892
QY 461 IYGDHINDTDFNKLNVK-DGDIIMDG-----MDQYKN-----G 496
DB 893 INGN-TNEKGDNLK-NIKDAELIQGNISQKEGNTLTSSDKVNITNQTIKAGVEGG 950
QY 497 FPDKYTMDGNVYLOT-----GY-----SDL-----NAKAVGYH 526
DB 951 RSDSEAEANALITITTKELKLAGDINISGFNAELITAKRGSULTIGNAGGADAKV--- 1008
QY 527 QFLYNVXPEVNIDPKGNTSIEYADGKSVENINDKRNNGFD--GEIOQHUYINGEY 584
DB 1009 --TFPKVK-----DSKIST-----DGHNVTLNSEVKTSGNSAGNDNSTGLTISAKDVT 1056
QY 585 SFNDIKQIIDTKLNIKIYVKOPARTTYKE-PILNKOTGVSSELKPHRYTYTONG----- 639
DB 1057 VNNVNTS--HKTINISAA---AGNVTKEGTTINATGSAE-----VTAONGTITKG 1102
QY 640 --KEMSTIVSEEDFI 653
DB 1103 NITSQNVTVTATENLV 1118

RESULT 13
US-09-200-650E-3
Sequence 3, Application US/09200650E
Patent No. 6680195
GENERAL INFORMATION:
APPLICANT: Patti, Joseph M.
APPLICANT: Foster, Timothy J.
APPLICANT: Hook, Magnus A.O.
APPLICANT: Eldholm, Delidre N.
APPLICANT: Perkins, Samuel L.
TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
FILE REFERENCE: P06283052/BAS
CURRENT APPLICATION NUMBER: US/09/200,650E
PRIOR FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: 60/066,815
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: 60/098,427
PRIOR FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 23
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 3
LENGTH: 930
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-200-650E-3

Query Match 4.6%; Score 186.5; DB 4; Length 930;
Best Local Similarity 20.8%; Pred. No. 0.0006;
Matches 175; Conservative 121; Mismatches 346; Indels 199; Gaps 42;

QY 3 GEIASKEKNLNGKESGLK-DTGVHHHQ---ENESIKESFTIDRNISTIRDFE 58
DB 56 GELNSKNEETAPSNKTKTKYVDSROLKDNQOTANADPKYTMDSALVKESSNMQSPQ 115
QY 59 NKDLKLIKKEFREVDFTSETGRMBEYDYKNDKNIIT-AYDDGTLEVEYEKLDIK 117
DB 116 NATANQSTTKTSNVTNNDKSSITVYSN-----TDKSNILQAKDVSTPKTTIKPRTLN 169
QY 118 SKIVGVLSPEKDGHEILIGKISNVSKNAKVY---GNNYKSIETKATKYDHSKTMTPDLYAN 177

```

Db      170 RMAVNTVAAPQO-----TNV--NDKVHESN---IDAIDKGHVNTQTKTEFWAT 215
Qy      178 INDI-----VDGLAFAGDMRLFYKNDQKAKIKIMPEKIKETKSEBPVSSYGV 229
Db      216 SSDVNLKLNKANYTIDDSVKEGD--TFYFKYGGYFRPESVRLP--SGTQNLVY--NAQGN 267
Qy      230 IELGEGDLSKNRP-----DNLTMESGKIYSDSEKQOYLKDNIIIRKGYALAKVTV 281
Db      268 IAKGIYDSTNTTNTTTFYTVVQYTVNRG-----SFEQVAFAPK-----RKATTDKTV 316
Qy      282 NPKETDMELENGVNGYKEDIKAKIOKANPNLRALSETTIYAD---SRNV-----EDGRST 331
Db      317 ---KHEVTLGNDYSEIIL--VDYGNKKAQPLISSTNYINNEIDSHMAYVAVNQPNQVTV 371
Qy      332 QSVLMSALDGF---NIIYQVTFPGMNDKGEAIDKGLVYDSSGLVAFGKDDKRYTGE 387
Db      372 KQTFYTNLTGYKFNPNAPKNFKI--YEVTONQFVD--SFTPDTSGL-----KQV-- 417
Qy      388 DKFNVEAIKEDGSMLEFIDTRPVNLSDKNYF-----NPSKSNKIYVRNBEFYLRGKI 439
Db      418 DQFDV-IYSNDMKTATVLDLKGQOTSNNKQYIIQQVAVPNSSSTDN-----GKI 464
Qy      440 ---SDKGFNWE--LAVNESVDNVIYIYGDLDHIDTRDENIKLVNKG--DLMWGM 489
Db      465 DYTLDTRTKYSMSNSYSNVNGSSTAN---GDQKCYNLGDYVWEDTNKDGKODANEKGI 520
Qy      490 KQY-----KANGRP--DKYTDMDGNVYLTQYSDLNKAVGVH-----YQFLVDNK 534
Db      521 KQYVYILKDSNGEELRRTTDENGKQYTGLSN-----GTYSVEFSTPAGYPTTANVG 574
Qy      535 PEVNIDPKNTS---LEVADG-----KSVFNINDK--RNGCFGEIOEQHIYINK 581
Db      575 TDADVDSGLTTTGVIKODNMNLTDSGFKTPKYSGLGVYWDYNSNDGKRDSTEKIKYV 634
Qy      582 EYTSFNDIKQIIDKTINIKIIVKDFPANTVVEFLNKDTGEVSEL--KPRHVTIION- 638
Db      635 KYTLQNEKEVIGTT-----ETDENGKTRFNNLDSGKKVIFEKRPAGILTQGTWT 684
Qy      639 -----GKMSSTVSEEDFILPYKGELEKGYQFDGMEISGFEKKDAGVYINLSMD 690
Db      685 TEDDKDADGGEVDVTTTDHDF-----TLNGYEETSDSDSDSDSDSDSDSDSDSDSD 737
Qy      691 TEIKVPFKIEKEKEENKPTFDSVKKQDPQVNSQLNESHKEDLQREHSQKSDYK 750
Db      738 S-----DSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 788
Qy      751 D 751
Db      789 D 789

RESULT 14
US-09-200-650E-5
; Sequence 5, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Eidiham, Delidre N1
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1315

```

```

; TYPE: PR1
; ORGANISM: Staphylococcus aureus
; US-09-200-650E-5

Query Match      4.6%; Score 186; DB 4; Length 1315;
Best Local Similarity 20.8%; Pred. No. 0.001;
Matches 188; Conservative 115; Mismatches 328; Indels 274; Gaps 51;

Qy      23 KQTTVEHHHNE-----ESTKESSFTIDNISITIRPEKDL---KK 64
Db      164 KSVVNVQPTNEENKVDADKTESSTLNKSDAITSDETLVDNN--SNSNNENNADILPFS 222
Qy      65 LKKKF-----REYVD--FTSETGRMEBYDYKDDKN--IAYVDGDTL 106
Db      223 TAPKLNTRMRIAAYQPSSTEAQVNDLITSITLTVVD---ADKNNKIYPADYLSL 277
Qy      107 EYETKLEIKS-----KIYGLSPSKDGHFEILAKISNVSQNAKAVYGGNNYK 154
Db      278 KQITVDDKVKSGDYFTIKYSDTVQVYG-LNPE-----DIKNI--GDIKDPNNGE 324
Qy      155 SLEIKATKYDFSKMTF--DLVANINDIVGLAFA-----GDKRLFYK-DN 198
Db      325 T--LATAKHDIANMLITTFYTDYVDRNSVQNGINYSIYMDADTLPUSKANDEFNVTIGN 382
Qy      199 DOKKAEIKIRMEKIKETKSEY-----PVSSYGN-----VIELGEGDLSKNK- 241
Db      383 TTTKTTANIQYPDYVNEKNSIGSAFTEYSHVGNKENGYKQTIYVNPSENSLTNAKL 442
Qy      242 -----PDNLTMESGKIYSD-SEKQOYLKDNIIIRKQY-----ALKV 278
Db      443 KYQAVHSSYFNNI-----QINKDVTDIKIYQVPKGYTLNKGVDVWTKELTVTNQYLQK 497
Qy      279 TTYNNGKTDMELENGVNGYKEDIKAKIOKANPNLRALSETTIYADSRNVEDGRSTQSLMSA 338
Db      498 IYGGNNNSAVID---FGNADSAVVMWNTKQ-----YNS--BSPTLVQATLSS 543
Qy      339 LDGFNIIRQVTFPKNDKGEA---IDKQNLV-IDSSK-----LVLFGK 379
Db      544 TGNKSVSTGNALGFNNQSGGAGQGEVYKIGNVWMDTNKGQVLEKGVGNTVTVFN 603
Qy      380 DKEVTGEDKFNVEAIKEDGSLFT---DTKPVNLSDKNY-FNPSKSNKIYVRNPEF 433
Db      604 NNNTKGE-----AVTKEDGSYLIPNLPNGDYRVEFSNLPKGEYVTPSKQG---NNBEL 654
Qy      434 YLRKGISDKGCFNWEIARVNESVVDNVIYGDLDHIDTRDENIKLVNKGSDIMDKWKQYK 493
Db      655 -----DSNGLSSVITVNGK--DN--LSADLGI-----YKPKNL--GDYV-W--EDTN 693
Qy      494 ANGFPDK-----VTDMDGNVYLTQYSDLNKAVGVHQPFL--YDNVKEPVNI-D 540
Db      694 KNGIQDQDEKGISGVTVTLKDENGNV-LKTVTTDADGK-----YKFTDLDNNGNYVEFTT 747
Qy      541 PKGNTSIEYADGKSVFNNIDRN---NGFDGEIOEQHIYNGKEYTSFNDIKQIIDKT 596
Db      748 PEGYTPTYVTSGDSDIEKDSNGLTGTVINGANMNLDSGFYKTPK---YNLGNVWEDT 803
Qy      597 LNIKIVVDFANTVTKERILNKDTGEVSELRP-HRVYTIIONGKMSSTIYSEEDFILP 655
Db      804 -----NKDGKQDSITEGIGSVYTLN-----ENGEVLD 832
Qy      656 VYKGELEKGYQFDGMEISGF--EGKKDAGVYINLSKQTFIKVPFKIEKEKEENKPTD 713
Db      833 TTKTDKDKGYQFTGLENGVYKVEFEPSPGYFT-----QVSGGTGDISNGSTTG 884
Qy      714 VSKKQDPQVNSQLNESHKEDLQREHSQKSDSTKD-----VTATVLDKNMISSKST 767
Db      885 VIKDKNDTIDSGFYKPTYNLGDYVWEDTNKNGVODKDEKGISGVTVTLKDENDKVLKTV 944
Qy      768 TNNPN 772
Db      945 TTDEN 949

```



```

RESULT 15
US-09-949-016-11433
; Sequence 11433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11433

```

```

Query Match 4.6%; Score 185; DB 4; Length 2733;
Beat Local Similarity 20.0%; Pred. No. 0.0032;
Matches 170; Conservative 150; Mismatches 338; Indels 194; Gaps 39;

```

```

QY 21 LKDDTGVHHQENEE--SIKESFTIDRNISTIDPENKDKLTKK---FREVDD 75
DB 1319 LKERIAGLEERKQKNEEFSQLENEKNITLSQIST-KQSELMLQEVTKMNLNQIOE 1377
QY 76 FTSETGKMEBYDYKYDD-----KGNIIAY-DDGTDLVEYTEKLDIKSIYIG 122
DB 1378 ELSRYTKLKETBEEDDLEERLNLQALMNGSIGNYCQDVTDAQIKNELLESEKMLKX 1437
QY 123 VLSPEKDHFEILGKISNVSKAKVYY-----GN-----NYKSIETKA 160
DB 1438 CVSELEERKQOLVKEKTYSEIRKRYLEKIQAOKEPENKSHAKLOELLKQOEYVQ 1497
QY 161 TKYD---FHSKTMTPDLVANDIYDGLAFAGDMLFVNDNDQKKAET-KIMPEKIKET 216
DB 1498 LQKDCIRIQEKSALERTYKALE-----FVOTESQKOLEITKENLAQAVEHR 1544
QY 217 KSEYPVSSYGVNIEIGEGDLSKNRPDLTKMESGKIYSDSEKQOYLKDNITLTK-- 272
DB 1545 KKAQAEIASFKYLDLDTQSEARVLAADNLKKELOSNKESYKSGMKQKQEDLERLEQA 1604
QY 273 -GYALKVTTYNPGKTDMLFGNGVYSEKEDIQKANKPNLRALSETTIVYADSRNVEDGRST 331
DB 1605 EEKHLKEKKMQEKDALAREKYLHEETIGEIO-----VTLNKQDKEVOO--- 1649
QY 332 QSVLMSALDGFNIIRYQVTFPGNDKGBAIDKQNLVTTDSKVLVFGKDDKEYTGEDKFN 391
DB 1650 ---LQENLDS-TVYQLAAFTKSM---SLQDDRDRVIDEAK---KMERKFSDALQSK 1696
QY 392 VEAL--KEDG-SMLEFIDTKPVNLMDKYNFNSK--SNKIY--VRNPEFYLRGKISDK- 442
DB 1697 EEEIRLKEDNCSVLKQDLROMSIHMEELKINISRLHDKQIWESKAQTEVQLOQKVCPTL 1756
QY 443 GGFNNEILAVNESVNNLYIGDLHDNTRDFNIKLVKDGDIIMDWGMKDYKANGFPDKYT 502
DB 1757 QGENKELL-----SOLEETR-----HLYHSQNEELAKLESELKSLKDQLT 1796
QY 503 DMD-----GNV--YLQTYGSDL-NAKAVGVHYQFLYDNVKKREV----- 537
DB 1797 DLSNLEKCKEKGKGNLBGIIIRQOEADIONSK-----PSYQLESTDLQASRELTSLHE 1849
QY 538 NIDPKGNTSIEYADSK--SVVENINDKRNNGFGEIOEQHIYINGKEYTSF--NDIKQI 592
DB 1850 EIMNKEQKTIISLSGKEBAIQVAIAELRQO-HDKKIKELLENLSQEEENIVLEENKKA 1908

```

```

QY 593 IDKTNINIKIVKDPARNTTVKEFIINKDQGEVSEIK--PHRYVTIQQNGKNSITIVSEE 650
DB 1909 VDKTQMLMETLTKTIKENIQQKAQDSEFKSMSSIQNDRDRIVDPYQOLEERHLSILEK 1968
QY 651 DFLIPVYGELEKGYQDFGWEISGPEG-----KKDAGVYINLSKDT 691
DB 1969 DOLIOEAALENKKKE-----EIRGLRSMDDLNSENALDDELIOYREDLNOVITI-KDS 2023
QY 692 FIKPVFK-KIEEKEEENKPTFDVSKKQNDPQVNHQUNES-----HRKEDLOREBSQK 745
DB 2024 QOKQLLEVQLOQNKLEKRYAKLEBKLESEBAN-EDLRSPNALQEEKQDLSKEIETSLK 2082
QY 746 ---SDSTQDVTA 754
DB 2083 VSIQLTQVTA 2094

```

Search completed: February 16, 2005, 19:20:41
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:19:57 ; Search time 141 Seconds
(without alignments)
1791.327 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGEIAESKFKVLGNKESG.....ATVLDKNISKSTNNPK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	773	13	US-10-067-385-8
2	4026	100.0	2119	10	US-09-769-744A-38
3	4026	100.0	2140	15	US-10-282-122A-73670
4	4026	100.0	2140	17	US-10-472-928-1180
5	615	15.3	117	9	US-09-765-872-68
6	227	5.6	1196	15	US-10-282-122A-52737
7	222	5.5	861	10	US-09-820-843A-109
8	218	5.4	1639	14	US-10-087-464-10
9	212	5.3	1141	15	US-10-282-122A-70251
10	207.5	5.2	4688	15	US-10-282-122A-76865
11	207	5.1	1191	15	US-10-282-122A-52048
12	203	5.0	1166	17	US-10-744-616-7
13	203	5.0	1881	14	US-10-032-585-7646

14	200.5	5.0	1786	9	US-09-742-096-3	Sequence 3, Appl1
15	200.5	5.0	1787	15	US-10-415-253-2	Sequence 2, Appl1
16	196.5	4.9	1184	15	US-10-282-122A-53254	Sequence 53254, A
17	196.5	4.9	5176	16	US-10-437-963-150986	Sequence 150986, A
18	195	4.8	1849	16	US-10-637-544-2	Sequence 2, Appl1
19	194	4.8	1178	15	US-10-282-122A-52434	Sequence 52434, A
20	193	4.8	1957	15	US-10-369-493-2070	Sequence 2070, A
21	193	4.8	903	15	US-10-282-122A-52328	Sequence 52328, A
22	193	4.8	909	15	US-10-282-122A-52109	Sequence 52109, A
23	192.5	4.8	708	17	US-10-472-928-3792	Sequence 3792, Ap
24	191.5	4.8	1156	15	US-10-369-493-11081	Sequence 43, Appl
25	190	4.7	1009	15	US-10-282-122A-43832	Sequence 43832, A
26	190	4.7	1788	15	US-10-282-122A-46664	Sequence 46664, A
27	189.5	4.7	841	9	US-09-815-242-5779	Sequence 5779, Ap
28	189.5	4.7	841	9	US-09-815-242-11751	Sequence 11751, A
29	189.5	4.7	1143	15	US-10-369-493-11081	Sequence 11081, A
30	189	4.7	1313	15	US-10-282-122A-76863	Sequence 76863, A
31	188	4.7	943	17	US-10-741-849-7309	Sequence 7309, Ap
32	187	4.6	1103	15	US-10-282-122A-76866	Sequence 76866, A
33	187	4.6	2265	15	US-10-282-122A-45123	Sequence 45123, A
34	186.5	4.6	930	17	US-10-744-616-3	Sequence 3, Appl1
35	186.5	4.6	1790	15	US-10-369-493-1586	Sequence 1586, Ap
36	186.5	4.6	1503	16	US-10-766-993-3	Sequence 3, Appl1
37	186	4.6	1315	17	US-10-744-616-5	Sequence 5, Appl1
38	186	4.6	1385	15	US-10-282-122A-44324	Sequence 44324, A
39	185	4.6	3225	16	US-10-408-765A-254	Sequence 254, App
40	185	4.6	5005	15	US-10-282-122A-76871	Sequence 76871, A
41	184.5	4.6	1946	15	US-10-282-122A-62947	Sequence 62947, A
42	184	4.6	1875	15	US-10-369-493-22285	Sequence 22285, A
43	183	4.5	2402	16	US-10-661-809-20	Sequence 20, Appl
44	182.5	4.5	882	15	US-10-282-122A-53247	Sequence 53247, A
45	182.5	4.5	923	16	US-10-771-931-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-10-067-385-8
Sequence 8, Application US/10067385
General Information:
Applicant: Adamou, John
Title of Invention: Streptococcus Pneumoniae Proteins and Vaccines
File Reference: 469201-589
Current Application Number: US/10/067,385
Current Filing Date: 2002-02-05
Prior Application Number: US/09/590,991
Prior Filing Date: 2000-06-09
Prior Application Number: US/60/138,453
Prior Filing Date: 1999-06-10
Number of Seq ID NOS: 8
Software: PatentIn Ver. 2.0
Seq ID NO 8
Length: 773
Type: PRT
Organism: Streptococcus pneumoniae
US-10-067-385-8

Query Match 100.0%; Score 4026; DB 13; Length 773;
Best Local Similarity 100.0%; Pred. No. 1.6e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLGEIAESKFKVLGNKESGLKDDTTGVHHHQNESIKESGFTIDNISTIRDFENK 60
1 KLGEIAESKFKVLGNKESGLKDDTTGVHHHQNESIKESGFTIDNISTIRDFENK 60
DB 61 DLKLIKKFRVDDPTSTFGKRMEDYDYKVDKNGIITAYVDGDTLDEYETKLDIKSKI 120
61 DLKLIKKFRVDDPTSTFGKRMEDYDYKVDKNGIITAYVDGDTLDEYETKLDIKSKI 120
DB 61 DLKLIKKFRVDDPTSTFGKRMEDYDYKVDKNGIITAYVDGDTLDEYETKLDIKSKI 120
121 YGVLSPSKDGHEIILGKISNVSKNAKVYGGNNYKSIKATKYDFHSKTMTPDLYANIND 180

```
Db      ||| 121 YGVLSPPSKDGHFEILIGKISNVSNAKVYGNVYSIEIKATYDPSKMTMDLVANIND 180
Qy      ||| 181 IVDGLAFAGDMRLFYKNDQKKAIEIKIRPEKIKETKSEYPIYVSYSYGVNIELEGGDLSEN 240
Db      ||| 181 IVDGLAFAGDMRLFYKNDQKKAIEIKIRPEKIKETKSEYPIYVSYSYGVNIELEGGDLSEN 240
Qy      ||| 241 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILKRGVALKVTYTPGKTDMLENGNGVYSKEDI 300
Db      ||| 241 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILKRGVALKVTYTPGKTDMLENGNGVYSKEDI 300
Qy      ||| 301 AKIOKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYOVFTPKANDKGEA 360
Db      ||| 301 AKIOKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYOVFTPKANDKGEA 360
Qy      ||| 361 IDKGNLVYDSSKLVLFKGDDEKEYTGEDKFNVEAIIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
Db      ||| 361 IDKGNLVYDSSKLVLFKGDDEKEYTGEDKFNVEAIIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
Qy      ||| 421 SKSNKIYYRNPEFYLRGKISDKGGFNWELRVNESVVDNYLLIYGDHIDNTDFNIIKLVNK 480
Db      ||| 421 SKSNKIYYRNPEFYLRGKISDKGGFNWELRVNESVVDNYLLIYGDHIDNTDFNIIKLVNK 480
Qy      ||| 481 DGDIMDMGKDYKANGFPDKYTDMDGNVYLTQGYSDILAKAVGVYQFLYDNVKEBVNID 540
Db      ||| 481 DGDIMDMGKDYKANGFPDKYTDMDGNVYLTQGYSDILAKAVGVYQFLYDNVKEBVNID 540
Qy      ||| 541 PKGNTSIEYADGKSVFVNINDKRNNGFDGEIOEHIIYNGKEYTSFNDIKQIIDTILNLIK 600
Db      ||| 541 PKGNTSIEYADGKSVFVNINDKRNNGFDGEIOEHIIYNGKEYTSFNDIKQIIDTILNLIK 600
Qy      ||| 601 IYVNDPANTTYKEFIILNKDTGSEVSELKPHRYVTYTIQNGKEMSSITVSEEDFILPVYKGE 660
Db      ||| 601 IYVNDPANTTYKEFIILNKDTGSEVSELKPHRYVTYTIQNGKEMSSITVSEEDFILPVYKGE 660
Qy      ||| 661 LEKGYQFDGWEISGEGKKDAGVYNLSKDTFIKPVFKKIEBKKEENKPTFDVSKKDN 720
Db      ||| 661 LEKGYQFDGWEISGEGKKDAGVYNLSKDTFIKPVFKKIEBKKEENKPTFDVSKKDN 720
Qy      ||| 721 PQVNSQLNESHKEDLQREHSHQSSTKDVATATVLDKNNISKSXTNNPNK 773
Db      ||| 721 PQVNSQLNESHKEDLQREHSHQSSTKDVATATVLDKNNISKSXTNNPNK 773

RESULT 2
US-09-769-744A-28
; Sequence 28, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hamblify, Sean B
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21122MO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: prt
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match      100.0%; Score 4026; DB 10; Length 2119;
Best Local Similarity 100.0%; Pred. No. 5.8e-216;
```

```
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      ||| 1 KLGEIAESKFNKILANGKESGLKQDTGYEHHHQBNEESIKEKSSFTIDRNISTITDFPNK 60
Db      ||| 1313 KLGEIAESKFNKILANGKESGLKQDTGYEHHHQBNEESIKEKSSFTIDRNISTITDFPNK 1372
Qy      ||| 61 DLKULIKKKFNEVDDFTSETGRMEEYDYKTDKKNIIAYDGDITLSEYETERLDEIKSKI 120
Db      ||| 1373 DLKULIKKKFNEVDDFTSETGRMEEYDYKTDKKNIIAYDGDITLSEYETERLDEIKSKI 1432
Qy      ||| 121 YGVLSPPSKDGHFEILIGKISNVSNAKVYGNVYSIEIKATYDPSHXTMTFEDLVANIND 180
Db      ||| 1433 YGVLSPPSKDGHFEILIGKISNVSNAKVYGNVYSIEIKATYDPSHXTMTFEDLVANIND 1492
Qy      ||| 181 IVDGLAFAGDMRLFYKNDQKKAIEIKIRPEKIKETKSEYPIYVSYSYGVNIELEGGDLSEN 240
Db      ||| 1493 IVDGLAFAGDMRLFYKNDQKKAIEIKIRPEKIKETKSEYPIYVSYSYGVNIELEGGDLSEN 1552
Qy      ||| 241 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILKRGVALKVTYTPGKTDMLENGNGVYSKEDI 300
Db      ||| 1553 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILKRGVALKVTYTPGKTDMLENGNGVYSKEDI 1612
Qy      ||| 301 AKIOKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYOVFTPKANDKGEA 360
Db      ||| 1613 AKIOKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYOVFTPKANDKGEA 1672
Qy      ||| 361 IDKGNLVYDSSKLVLFKGDDEKEYTGEDKFNVEAIIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
Db      ||| 1673 IDKGNLVYDSSKLVLFKGDDEKEYTGEDKFNVEAIIKEDGSMFLFIDTKPVNLSMDKNYFNP 1732
Qy      ||| 421 SKSNKIYYRNPEFYLRGKISDKGGFNWELRVNESVVDNYLLIYGDHIDNTDFNIIKLVNK 480
Db      ||| 1733 SKSNKIYYRNPEFYLRGKISDKGGFNWELRVNESVVDNYLLIYGDHIDNTDFNIIKLVNK 1792
Qy      ||| 481 DGDIMDMGKDYKANGFPDKYTDMDGNVYLTQGYSDILAKAVGVYQFLYDNVKEBVNID 540
Db      ||| 1793 DGDIMDMGKDYKANGFPDKYTDMDGNVYLTQGYSDILAKAVGVYQFLYDNVKEBVNID 1852
Qy      ||| 541 PKGNTSIEYADGKSVFVNINDKRNNGFDGEIOEHIIYNGKEYTSFNDIKQIIDTILNLIK 600
Db      ||| 1853 PKGNTSIEYADGKSVFVNINDKRNNGFDGEIOEHIIYNGKEYTSFNDIKQIIDTILNLIK 1912
Qy      ||| 601 IYVNDPANTTYKEFIILNKDTGSEVSELKPHRYVTYTIQNGKEMSSITVSEEDFILPVYKGE 660
Db      ||| 1913 IYVNDPANTTYKEFIILNKDTGSEVSELKPHRYVTYTIQNGKEMSSITVSEEDFILPVYKGE 1972
Qy      ||| 661 LEKGYQFDGWEISGEGKKDAGVYNLSKDTFIKPVFKKIEBKKEENKPTFDVSKKDN 720
Db      ||| 1973 LEKGYQFDGWEISGEGKKDAGVYNLSKDTFIKPVFKKIEBKKEENKPTFDVSKKDN 2032
Qy      ||| 721 PQVNSQLNESHKEDLQREHSHQSSTKDVATATVLDKNNISKSXTNNPNK 773
Db      ||| 2033 PQVNSQLNESHKEDLQREHSHQSSTKDVATATVLDKNNISKSXTNNPNK 2085

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykling, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
```

```
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent version 3.1
SEQ ID NO: 73670
LENGTH: 2140
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670
```

Query Match 100.0%; Score 4026; DB 15; Length 2140;

Best Local Similarity 100.0%; Pred. No. 5.9e-216; Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
1 KGEIAESKFKNLGNGKESLKKDVTGVHHQENESIKESSTFIDNISTIRDFENK 60
1334 KGEIAESKFKNLGNGKESLKKDVTGVHHQENESIKESSTFIDNISTIRDFENK 1393
61 DLKLIKKKFRVDDFTSTGKRMEDYDYKDDKGNIIAYDGTDLLEYTEKLDKSKXI 120
1394 DLKLIKKKFRVDDFTSTGKRMEDYDYKDDKGNIIAYDGTDLLEYTEKLDKSKXI 1453
121 YGVLSPSKDGHEIILGKISNVSKNAKVYGNVYKSIIEIKATKYDFHSKTMTEFDLYANIND 180
1454 YGVLSPSKDGHEIILGKISNVSKNAKVYGNVYKSIIEIKATKYDFHSKTMTEFDLYANIND 1513
181 IYDGLAFADDMRLFYVNDNQKAEIKIRMEPEIKETKSEYVSSYGVNIELEGEDLSKN 240
1514 IYDGLAFADDMRLFYVNDNQKAEIKIRMEPEIKETKSEYVSSYGVNIELEGEDLSKN 1573
241 KPDNLTMSGSKTYSSEKQOYLKDNITLRKGYALKVTTYPNGKTMLEGNVYSKEDI 300
1574 KPDNLTMSGSKTYSSEKQOYLKDNITLRKGYALKVTTYPNGKTMLEGNVYSKEDI 1633
301 AKIQKPNLRLSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVTFPRMNDKGBA 360
1634 AKIQKPNLRLSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVTFPRMNDKGBA 1693
361 IDKGNLVTDSKLVLFGKDDKEYTGEDKFNVEAIKEDGSMFLPITKPNVLSMDKNYFNP 420
1694 IDKGNLVTDSKLVLFGKDDKEYTGEDKFNVEAIKEDGSMFLPITKPNVLSMDKNYFNP 1753
421 SKSNKTYVNPPEYLLGKISDKGFWELRVNESVVDNLIYGDHLIDNTRDNIKLANY 480
1754 SKSNKTYVNPPEYLLGKISDKGFWELRVNESVVDNLIYGDHLIDNTRDNIKLANY 1813
481 DGDIMGMKDYKANGFPDKVTMDGNVYLTQGSYDLNKAAGVNHQPLYDNVKNPEVND 540
1814 DGDIMGMKDYKANGFPDKVTMDGNVYLTQGSYDLNKAAGVNHQPLYDNVKNPEVND 1873
541 PKGNISIEYADGKSVFENIDKRNNGFDEIOEHYINGKEYTSFNDIKQIIDTKTANIK 600
```

```
1874 PKGNISIEYADGKSVFENIDKRNNGFDEIOEHYINGKEYTSFNDIKQIIDTKTANIK 1933
601 IYVNDPAVTTVKEFLINDQGEVSELKPHRTVITIQNKENSSTVSEDETLIPYKGE 660
1934 IYVNDPAVTTVKEFLINDQGEVSELKPHRTVITIQNKENSSTVSEDETLIPYKGE 1993
661 LEKGYPDGMELISGEGKKAGVYNLSKDTFIKPFKIEKEKEENKPTFDVSKKON 720
1994 LEKGYPDGMELISGEGKKAGVYNLSKDTFIKPFKIEKEKEENKPTFDVSKKON 2053
721 POWNHSQLNESHKEDLOREHSHQSDSTKDVATVLDNNISSTXTNNPNK 773
2054 POWNHSQLNESHKEDLOREHSHQSDSTKDVATVLDNNISSTXTNNPNK 2106
```

RESULT 4

US-10-472-928-1180

Sequence 1180, Application US/10472928

Publication No. US20050020813A1

GENERAL INFORMATION:

APPLICANT: CHIRON SPA

APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH

TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS

FILE REFERENCE: P026926WO

CURRENT APPLICATION NUMBER: US/10/472,928

CURRENT FILING DATE: 2003-09-26

PRIOR APPLICATION NUMBER: GB-0107658.7

PRIOR FILING DATE: 2001-03-27

NUMBER OF SEQ ID NOS: 4979

SOFTWARE: SeqMin99, version 1.03

SEQ ID NO: 1180

LENGTH: 2140

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

FEATURES:

OTHER INFORMATION: serine protease, subtilase family

OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPYTG)

OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)

US-10-472-928-1180

Query Match 100.0%; Score 4026; DB 17; Length 2140;

Best Local Similarity 100.0%; Pred. No. 5.9e-216; Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
1 KGEIAESKFKNLGNGKESLKKDVTGVHHQENESIKESSTFIDNISTIRDFENK 60
1334 KGEIAESKFKNLGNGKESLKKDVTGVHHQENESIKESSTFIDNISTIRDFENK 1393
61 DLKLIKKKFRVDDFTSTGKRMEDYDYKDDKGNIIAYDGTDLLEYTEKLDKSKXI 120
1394 DLKLIKKKFRVDDFTSTGKRMEDYDYKDDKGNIIAYDGTDLLEYTEKLDKSKXI 1453
121 YGVLSPSKDGHEIILGKISNVSKNAKVYGNVYKSIIEIKATKYDFHSKTMTEFDLYANIND 180
1454 YGVLSPSKDGHEIILGKISNVSKNAKVYGNVYKSIIEIKATKYDFHSKTMTEFDLYANIND 1513
181 IYDGLAFADDMRLFYVNDNQKAEIKIRMEPEIKETKSEYVSSYGVNIELEGEDLSKN 240
1514 IYDGLAFADDMRLFYVNDNQKAEIKIRMEPEIKETKSEYVSSYGVNIELEGEDLSKN 1573
241 KPDNLTMSGSKTYSSEKQOYLKDNITLRKGYALKVTTYPNGKTMLEGNVYSKEDI 300
1574 KPDNLTMSGSKTYSSEKQOYLKDNITLRKGYALKVTTYPNGKTMLEGNVYSKEDI 1633
301 AKIQKPNLRLSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVTFPRMNDKGBA 360
1634 AKIQKPNLRLSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVTFPRMNDKGBA 1693
361 IDKGNLVTDSKLVLFGKDDKEYTGEDKFNVEAIKEDGSMFLPITKPNVLSMDKNYFNP 420
1694 IDKGNLVTDSKLVLFGKDDKEYTGEDKFNVEAIKEDGSMFLPITKPNVLSMDKNYFNP 1753
```

QY 421 SKSNKIYVNPBEPYLRGKISDKGPFNMEIRVNESVVDNLYIYGDHIDNTRDNPNTKLNK 480
DB 1754 SSNNKIYVNPBEPYLRGKISDKGPFNMEIRVNESVVDNLYIYGDHIDNTRDNPNTKLNK 1813
QY 481 DGDINDMGKDKYKANGFPDKYTDMDGNVYLQGYSPDLNKAAGVHYQFLYDVKPEVND 540
DB 1814 DGDINDMGKDKYKANGFPDKYTDMDGNVYLQGYSPDLNKAAGVHYQFLYDVKPEVND 1873
QY 541 PKGNTSIEYADGKSVFENINDKRNNGFDEIOEHYIYNGKEYTSFNDIKQIIDKTLNIX 600
DB 1874 PKGNTSIEYADGKSVFENINDKRNNGFDEIOEHYIYNGKEYTSFNDIKQIIDKTLNIX 1933
QY 601 IYVKDPAFTTVEKFLNKDNGEVSLEKPHRYVTIYQNGKEMSTTVSEDFILPYKGE 660
DB 1934 IYVKDPAFTTVEKFLNKDNGEVSLEKPHRYVTIYQNGKEMSTTVSEDFILPYKGE 1993
QY 661 LEKGQFQDMEISGEGKKDAGVYINLSKDTFIKPVFKKIEEKEENKPTFDVSKKDN 720
DB 1994 LEKGQFQDMEISGEGKKDAGVYINLSKDTFIKPVFKKIEEKEENKPTFDVSKKDN 2053
QY 721 POWNHSQLNESHKEDLQREHHSQKSDSTKDYATATVLDKNNISSKSTNNPNK 773
DB 2054 POWNHSQLNESHKEDLQREHHSQKSDSTKDYATATVLDKNNISSKSTNNPNK 2106

RESULT 5

US-09-765-272-68
Sequence 68, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68

Query Match 15.3%; Score 615; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.4e-27;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQPDGWSISGEGKKDAGVYINLSKDTFIKPVFKKIEEKEENKPTFDVSK 716
DB 1 YKGELEKGYQPDGWSISGEGKKDAGVYINLSKDTFIKPVFKKIEEKEENKPTFDVSK 60
QY 717 KKDNPVNHSQLNESHKEDLQREHHSQKSDSTKDYATATVLDKNNISSKSTNNPNK 773
DB 61 KKDNPVNHSQLNESHKEDLQREHHSQKSDSTKDYATATVLDKNNISSKSTNNPNK 117

RESULT 6

US-10-282-122A-52737
Sequence 52737, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 52737
LENGTH: 1196
TYPE: prt
ORGANISM: Clostridium botulinum
US-10-282-122A-52737

Query Match 5.6%; Score 227; DB 15; Length 1196;
Best Local Similarity 20.2%; Pred. No. 0.00029;
Matches 185; Conservative 145; Mismatches 260; Indels 328; Gaps 46;

QY 1 KLGEIAESKPFNKLNGKESG-----SLKDTTGVHHHGHENESR 39
DB 123 RLKQIOE-LFMDYTGKGGYSLIGGKIEAVLSGKPEERRSLLEBAAGIVKPKKKEAD 181
QY 40 KEKSFITDKNISTRDEN--KDIKKIKKKFREVDV--TSETGKME----- 85
DB 182 KKLNSN--TEONLIRIKDILNTYEEEMPELKESEKAKKFLNLSEELKKEVNVVMYISDK 239
QY 86 -EYDKYVDKNITIAVYDGTD-LVEYTEKLDEISIKYIGVSPSDDGHFELIGKISNYSK 143

```

Db      240 IERDLK-NISSNMLSGENIDMLKNEKSOYEKIEISKF-----NEKLELDK--NNSR 288
Qy      144 NAKVYGNPNKSEIKATKYDHFSTKMTEDLVANINDIYDGLAFADMLFVKDNOQKA 203
Db      289 NKEEYNNKDKNKDIE-----NENALKEKIKNLDK-----NIVKENTLKTN 331
Qy      204 EIKIR-----MPEKIKETKSEYPPV----- 223
Db      332 EERLQYKLEKKEALKEKINKLKEESYLRREIEKEDNVNPNKLEKEKLIKLSER 391
Qy      224 -----SYGNVIELEGDL--SKNKPMLTMESEKISY-----DSKQOYLK 265
Db      392 IEILNSTMLRNEISVMEWEIVLLENKLDINKSCSYISINIKITKEDIKEIKIK 451
Qy      266 DNILL-----RKGYALKVTTNPGKIDMLEGNVYSKED-----IAKIOK--AN 307
Db      452 ENILLENLKNKERSNISSIKISLNNKEK-KLKEKNAAYSRLKANYHMLSNLEKHYEGYN 510
Qy      308 PNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPKMDKGAIDKXGNL 367
Db      511 RSVKTLMEHVSYKVNINIKG-----GEVLGDIIKVKKELETAMEIALGAI-----SNV 560
Qy      368 VIDSCKVLVFGKDDKEYTEGDKFNVEAIKEDGSMFLIDTKPVNLSDMKRYFNP--SKSN 424
Db      561 IIEDE-----NKAKILINYLKCK-----SLGRATFLTLTTIOGR 594
Qy      425 KIYVRNPEFYLRKGISDKGSPNMLRVNESVVDNYLIYGDHLIDNTRDFNI--KLNVKOG 482
Db      595 KAKIN-----VREDFG--LGIASDLIY--DVKFSNIIDVYLGRTLVAKD- 637
Qy      483 DIMDMGMDKYKANGFPDKVLTDMGNYVLTQVYSDLNAKAVGVHGYOFLYDNVKEPVNIDPK 542
Db      638 --MDSALKJLAKLNTYSFKLVTLEGEV-----INPG 665
Qy      543 GNT--SIEYADKSVVFNINDKRNNGFDGEIOEHIYINGKAYTSPNDIKIIDTLNLI 599
Db      666 GSIJGSGIKHRAGSSI--ISRRK-----EIRE-----TKKLEBELTKN- 700
Qy      600 KIVVDFANTVYKEPILNKDGTGEVSELKPHRVLTVTIONGKEMSSITVSEEDTLFVYK 659
Db      701 -----TIEEPNKN-----ILENNKIKIT--LDEENLNT--KD 728
Qy      660 ELEKGYOPDGEWISGFEKKDA-----GYVINLSKD--TFIKVFPKKIEKKEEENK 709
Db      729 EI-----YANNIEITKMGKMLAIKEDTERLRSSLSINISREIKLTKOKIDIEENINISOK 784
Qy      710 PTFVSKKQDPVNSQALNESHKEDLQREHSHQSQSDSK--DVTATVLDKNNISK 765
Db      785 QLEELKDKD--LNNHNDICE--EDFLQNEENVAKIKDKLIEYKIEKAKLDEMVLVSIK 839
Qy      766 -----STNPNK 773
Db      840 KELYSMDTNTITLNNENK 857

RESULT 7
US-09-820-843A-109
; Sequence 109, Application US/09820843A
; Publication No. US2003003963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820, 843A
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 109
; LENGTH: 861
; TYPE: PRF
; ORGANISM: Plasmodium falciparum
; FEATURE:

```

```

; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|4493994
US-09-820-843A-109

Query Match      5.5%; Score 222; DB 10; Length 861;
Best local Similarity 19.3%; Pred. No. 0.00037;
Matches 162; Conservative 136; Mismatches 324; Indels 216; Gaps 37;

Qy      2 LGEIASESKPKNTANG-----KESGLKQDTTGYEHHN--QENESIKESKSFIT 47
Db      154 LNNINGKVFKNLDNDQCNVLPNTNLYIDKEG--KMLTGKEHYNAASNEYNHNNKNTNYY 211
Qy      48 DNISITIRDFEKDL-----KLLIKKKFREVDDFSEYKMEEDYKDYKDDGNTI 98
Db      212 NNNSTNNNNFPCNNNYNDNNYNNNSNNGKNTKERSLNTYK--KEHDVYDEYNKGNIR 268
Qy      99 AYD-----DGTDLLEYETEKLDIEIKSKIYGLVSPSKDGHFEI--LQKISNVSRNAYYGN 151
Db      269 KNDSEKYNWNPPLHY-----SKKNVDTFTIGDIKKAAYAKNNEKGN 310
Qy      152 NTKSIEIKATKYDHFSTKMTEDLVANINDIYDGLAFADMLFVKDNOQKAIEIKR 208
Db      311 KTNMNM-----HDNNSN--NSNNVNNNNNNNSNNYNNIFKONDENTL----- 351
Qy      209 MEKIKETKSEYPSYGNVIELEGD--DLSK--NKPNMLTMESEKISYDSEKQOYLL 264
Db      352 -----TNSNPAKFNKNNNNNANVENDTIITLANK-----NSQHSIDKNNNNN 396
Qy      265 KNNILRKGYALKVTTNPGKIDMLEGNVYSKEDIKIQKANPN--LRALSETTIYAD 321
Db      397 GNNII-----NNSNNKNNIIFQSGNRNRYENWVYNNNNNNNNIISNNKGEASFNTD 447
Qy      322 SRNVEDGRSTQSVLMSALDGFNIIRYQVTFPKMDKG-EAIDKGNLVTDSCKVLVFGKD 380
Db      448 NINTNSGREEEKISNTVAE--LLMKQISIMIKERNGLDVLREKNT-----FGPL 494
Qy      381 DKEYGEDFNVEAIEEDGSMFLIDTKPVNLSDMKRYFNPSKNTKYVRNPEFYL----- 435
Db      495 DNNYQYGSNN-----NSSLEKNNM--KENDIYSKEASKRINDIR 533
Qy      436 ---RGKISDKGSPNMLRVNESVVDNYLIYGDHLIDNTRDFNIKLVNKGDMGMDKMD 491
Db      534 TINSNGLVSGE-----SLVNSQSVLNNNNNNYNNNSNRRNKNQNNNNNNNNM----- 582
Qy      492 YKANGFPDKVLTDMGNYVLTQVYSDLNAK--AVGYHYQFLYDNVKEPVNIDPKNTSIE 548
Db      583 -----NNMNNNSNNNNNNNNNNYNNKNNHKKYHSGMDVYTKKIF--INNSGNNNGNNNSNS 633
Qy      549 YADGKSVFNINDKRNNGFDGEIOEHIYINGKAYTSPNDIKIIDK----- 595
Db      634 NSNNNVEHYNNNNKN--FKNKLNNYHNLPDNKNMANNNTYNNINKNLMSIENPEPSL 691
Qy      596 TLNLIKIVDFAFN--TTVKEFILNKDGTGEVSELKPHRVLTVTIONGKEMSSITVSEED 651
Db      692 SFNNSDINKNNQGNINITPIINSILRLD-NEVDVHNNNISSEINQNAK--VSNVLDLSK 748
Qy      652 FILPVYKGLKGYOPDGEWISGFEKKDAQVYINLSKTFK-----PVPKIEKK 704
Db      749 SLLKSKSGGNNNNYNIIP--KNFNNNNN-----NNNSKFTINYSQOYYPSHQOQOQH 799
Qy      705 EENKKTPTVSKKQDPVNSQALNESHKEDLQREHSHQSQSDSK--SDSTQVATVLDK 759
Db      800 QQQQQQ--QQQQLIOTIOTINSHLNDFNKKKKRKKERYPMKTPPEFGTTNEMVAEK 855

RESULT 8
US-10-087-464-10
; Sequence 10, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Christl, Athar

```

```

; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: SI237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; US-10-087-464-10

Query Match      5.4%; Score 218; DB 14; Length 1639;
Best Local Similarity 18.6%; Pred. No. 0.0014;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

QY 23 KDTTG-VEHHQENESIK-----EKSFTIDRNISTIDPENKDLKLIKRRBYVD 75
DB 232 KDNVGMEDYIKKAKKTIENINELIBESKTIIDKKNATKEE-----KKKLYOAY 283
QY 76 FTSETGKMEEDYKYDKGNIAYDGTDLSETEKLDIISKIYGVLSPEKDHFEIL 135
DB 284 DLSIYNKQEE-----AHNLISV-----LE-KRIDTLK-----KNEIKELL 319
QY 136 GKISIVSKAKKAYYGNKYSIKATKYDFH-----SKMTFPLYNINDIVDGLAFA 188
DB 320 DKINEIKKPPANSNTPTLADKNKKI EENKEIKELAKTIKTFIDSLFTPL----- 373
QY 189 GDMRLFVNDNDCKKAIKIRMEKIKETSEYV-----YVSSYGVN-----IEGE-GDLSK 239
DB 374 -ELEYYLEKKN-KNIDISAKVETKESTEPENFPGVTTPYLSINDINNALNEANSGLDL- 430
QY 240 NKPDNLTYMGESGIYSDESKOYL--LKDNILL-----RKGYALKVTTYNPGKTDL 289
DB 431 -NPDYTKPEPSKNITVDNERKKFINIEKIKIEKKKIESDKSYEDRSKSLN---DIT 485
QY 220 EG-----NGVYSKE-----DIKIQKA-----NPMRLSETTYIADSANVEDGSTOSV 334
DB 486 KEYEKLMEIYOSKNNNNIDLTNFERKMGKRYSYVEKLTHTNTASYSNSGMLKLTK 545
QY 335 IMSALDGF--NII-----RYOVFTPKNDKGEAIDKGNLVTSKSLVLFSGDKDEY 385
DB 546 ALKWMEDYSLRIVVEKELKYYKNIISKIENIEITLVE--NIKDEEQLF-----EKKIT 598
QY 386 GEDKENVEAIKEDGSMFLIDTRPVNLSDMKYFNPSKSNKIYVRNPEFYLRGKISDKGPF 445
DB 599 KDNKRDEKILEVSDIVKVOQKVL--MNKIDELKQTQLLKAVE--LKNINIVPSY 653
QY 446 NWEELRVNESVDNNTIYGDHLIDNTRDRIK-----NVK----- 480
DB 654 KOENKOE--PYLILVILKEIDKLKVPKVESLINEKKNIKTEGOSDNESEPTGEI 709
QY 481 -----DGDIMMGAKDYK-----ANGFPDKV 501
DB 710 TGOATTKGQQAAGSALBEGDSVAQAQOEKQAPPVVPRVPEAKAQPPTPPRPAVNNKTEV 769
QY 502 TDMGNVYIKQGYSDLANA-----KAVGVHYQFLYNNVREVINIDPKNGTSLIEYADGKSV 556
DB 770 SKLD--YLEKLYEFLNTSYICHKTYLVSHSTNMEKILKQYKITEESKSSCDPLDL 826
QY 557 FNIN-----DKRNNGPDG--EIOEGHIYINGKEYTSFNDIKQIIDLTLNIKIVY 603
DB 827 FNIONNIPWYSWFMPSLNNLSQLFMEIYIEKEMVONCLVKDKNDIKMLLEAKKVSISY 886
QY 604 KDFANNTIYKEFILNKDGEVSELKPHRYTATYIONGKEMSTIYSEEDPILFVYGELEK 663
DB 887 KTLSSSQPLSLTITQDKREYVANDTSHSTNLSKLIFENILS-----LGNKN 937

```

```

QY 664 GYQPDGWEISGEFGK-----KDAGVINLSKDTFIKVPFKKIEEKEENKPTF--D 713
DB 928 IYQ-----ELIGKSSSENFYEKILKSDPTFNYSFTNFKSKRADINSINDESKKKLEED 993
QY 714 VSKKDNQVNVHNSQLNESHKREDLQREHSQKSDSTKVYATVLDKNNISK-STNNP 771
DB 994 INKLKKTITQLSFDLNKKYKLEBLFDPKKTGVGKYKWOIKULTTLKEQLSKSLNPNP 1052

RESULT 9
US-10-282-122A-70251
; Sequence 70251, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haebelbeck, Robert
; APPLICANT: Olsen, Karl
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70251
; LENGTH: 1141
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-10-282-122A-70251

Query Match      5.3%; Score 212; DB 15; Length 1141;
Best Local Similarity 20.9%; Pred. No. 0.0019;
Matches 182; Conservative 116; Mismatches 336; Indels 238; Gaps 44;

QY 33 GENESITSE--KSFTIDRNISTIDPENKDLKLIKRR-----EYDDPTS 78
DB 221 KNPEKLELYANDNNTDRSTKPVATFSAAPKRLNAKMFPAVAQAAPAAVANNVNDLIT 280
QY 79 ETGKMEEDYKYDKGNIAYDGTDLSEYTE-----KLDIISKIYGVLS 125
DB 281 VTKQIK-----VGGKDNVAAAHGKQDIEYDTEFTIDNKVKKSGDMTINTNYKNVPSLT 336
QY 126 PSKD-----GHEIILGKISNVSKNAK---VYGNKYKSIKATKYDFHSGKTMTFDL 174

```

Ds	337	DKNDIDIDDPGSEVIAKGTFPKATQIYVTFYDYDKEDIKARLTIVSY-----IDK	390
Oy	175	YANINDYDGLAPA-----GD---WLFVK-DNDOKAEIKRM-	209
Ds	391	QAVPNETSINLTFATGAKETSGQNSVDYODPMWGHGDSNIQISIFTKLDENQOTIEQIUYVN	450
Oy	210	PEKIKETSEY-----PYVSYGNGVIELSGD--LSNKKRPNLTKM-----ESGKIYSD	256
Ds	451	PLKKTATNTKVDIAGSQVDYGN-ITLGGNSGTIIIDNTETELKYKVNPNQOLPQSNRIYDF	509
Oy	257	SEKQOYLTK-DNIIILRKGYALKVTTYNPGKTDLLEGNGVYSKEDIKIKRANPNLRALSE	315
Ds	510	SOYEDVTSQFDN---KKSFSSNNVATIDFG-----DINSAYI-----IKVSK	548
Oy	316	TTIYADSR-NVEDGRSTQGLMSALDGFNIIRYQVTFPKKNDG-----	358
Ds	549	YTPSTDGDELIDAGTSMRTT--DKGYVNVAGSYNFIATSNIDTGGGDTGVKPEKLYKIG	606
Oy	359	---EALDDGMLVYDSSKLVLPGKDDKXYTGSGDKRNVKAIKDGSMFLDTYEPNALSM	414
Ds	607	DYVWEDVDKDG-----VQGTDSKKEPMANVL-VTLTYPDGT-----TKSVRTDAN	650
Oy	415	KNY-FNPSKSNKIYV---RNPEFYLRGKIS-----DKGFMELRVNESVVDNVLIVG	463
Ds	651	GHYEEGGLKDGETYVYKFEFTPAGYLFYKNGTTDGEKDSNGSSITYKINCK-----D	702
Oy	464	DLHID-----NTRDENIKLVNKG--DINDMGKQYKANGFPDKVYDMDGNVYLQT	512
Ds	703	DMSLDTGFYKEPKVNLG DYVWEDVNTKDG IQDANEPGIDKVYV-----LKDSTGKV-IGT	756
Oy	513	GYSIDLAKAVGVHYOPL-YDNKKPEVNID-PKG-----NTSIEVAD-GKSVVFENINDK	562
Ds	757	TTTDSAGK-----YKFTDLDNGNYTVEFERPAGYTPYKNTTAEDDXSNMLTTGGYIKDA	811
Oy	563	RNNGPDELQEGHIYV-----NGKEYTSFNDIKOIIDKTLNIK-IYVKDFABNT	610
Ds	812	DWMTLDSGCKYKPKKISLGDYVWYDSNKKDGKODSTEGIKIDKVKYVTLNKEKEVIGTTKDE	871
Oy	611	TYKEFI LNKTGSEVSEL-KPHRYVTYIION-----GKEMSTYVSEDEPILPYKKG	659
Ds	872	NGKYPFMDNDSGKYKVIPEKPAGLTQTVNNTTEDDKDAOGGEVDVITITHHDFI-----	925
Oy	660	ELEKGYQDFGMEISGCEGKKDAGYVNLSDKTFIKYFPFKIIEKKBEENKPTPVSKKCO	719
Ds	926	-LDNGYFEEHDTSD	975
Oy	720	NPQVHNSQLNESHKEDDLQREHSQKSDSTKD	751
Ds	976	SD	1007
<p>RESULT 10</p> <p>US-10-282-122A-76865</p> <p>; Sequence 76865, Application US/10282122A</p> <p>; Publication No. US20040029129A1</p> <p>GENERAL INFORMATION:</p> <p>; APPLICANT: Wang, Llangau</p> <p>; APPLICANT: Zamudio, Carlos</p> <p>; APPLICANT: Malone, Cheryl</p> <p>; APPLICANT: Habelbeck, Robert</p> <p>; APPLICANT: Ohlsen, Karl</p> <p>; APPLICANT: Zyskind, Judith</p> <p>; APPLICANT: Wall, Daniel</p> <p>; APPLICANT: Trawick, John</p> <p>; APPLICANT: Carr, Grant</p> <p>; APPLICANT: Yamamoto, Robert</p> <p>; APPLICANT: Forsyth, R.</p> <p>; APPLICANT: Xu, H.</p> <p>; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms</p> <p>; FILE REFERENCE: ELIIRA 034A</p> <p>; CURRENT APPLICATION NUMBER: US/10/282, 122A</p> <p>; CURRENT FILING DATE: 2003-02-20</p>			

```

      PRIOR APPLICATION NUMBER: 60/191,078
      PRIOR FILING DATE: 2000-03-21
      PRIOR APPLICATION NUMBER: 60/206,848
      PRIOR FILING DATE: 2000-05-23
      PRIOR APPLICATION NUMBER: 60/207,727
      PRIOR FILING DATE: 2000-05-26
      PRIOR APPLICATION NUMBER: 60/230,335
      PRIOR FILING DATE: 2000-09-06
      PRIOR APPLICATION NUMBER: 60/230,347
      PRIOR FILING DATE: 2000-09-09
      PRIOR APPLICATION NUMBER: 60/242,578
      PRIOR FILING DATE: 2000-10-23
      PRIOR APPLICATION NUMBER: 60/253,625
      PRIOR FILING DATE: 2000-11-27
      PRIOR APPLICATION NUMBER: 60/257,931
      PRIOR FILING DATE: 2000-12-22
      PRIOR APPLICATION NUMBER: 60/267,636
      PRIOR FILING DATE: 2001-02-09
      PRIOR APPLICATION NUMBER: 60/269,308
      PRIOR FILING DATE: 2001-02-16
      Remaining Prior Application data removed - See File Wrapper or PALM.
      NUMBER OF SEQ ID NOS: 78614
      SOFTWARE: PatentIn version 3.1
      SEQ ID NO: 76865
      LENGTH: 4688
      TYPE: PRT
      ORGANISM: Ureaplasma urealyticum
      US-10-282-122A-76865

Query Match      5.2%; Score 207.5; DB 15; Length 4688;
Best Local Similarity 19.6%; Pred. No. 0.02; Indels 257; Gaps 47
Matches 179; Conservative 142; Mismatches 342;

21 LKQDTTGVHHHOENESIKESSTIDRNISTIDFENKQIKLKKKPREVDFTST 80
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db INKQVGV--FVDENQGVKIKTQ--IDYAKF--EPTKYLPKTGKTKLNIIDLNNI 131
      78

81 GKREERYDYKTDKGNIT-----AYDDGTLEFETEKULDEI-----KSKY 121
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db VRENEK--LKEQKLEVLKPVPTWYVEDQKNEKHQIKLVKNKILAKVKEVNQNEVY 189
      132

122 GVLSPSKDGHHELLGKI---SNVSNKAVVYGGNNVKSIEIKATKDPHSKMTFPLYANI 178
      122

190 SI-----EAHGLQCKIDIDITSLSKQGVVY--LKDLILK-----DSNPILT--KVNV 233
      190

179 NDIVDLAFAADMELFVNDQKTAELIKIRMEPIK-----ETKSEY--PYVSS 225
      179

234 NDIBE--QFKKIDITHYIDNENGLV-LKVDLTNKTQNALPTAVPKDENKKEYILPALSV 290
      234

226 YGVNIELEGSDLSKKNPDLTKMESGKIYDSEKQOYLLKONIILKGYALKVTTYNPCK 285
      226

291 KNNV-----GYENFKKLPKLNKHYELNRIVNNNELNDVLPVNNELMLE--HKVSVNKPTRAK 344
      291

286 TDMLEGNGVYSKEKDIATQKXNP--NLRALSETTYIADSRVNEGCRSTQVLSMALDSF 342
      286

345 VNFIDNKKVY--EVLGQVQKNTPELTLEEDLNHQYKIKNAIKDEKGRAVFDI--SSLGDN 401
      345

343 NIIRYQVETFGANRKGSAID-----KQGNL-----VTDSKVLVPGKDBE 383
      343

402 NL--YEVIGIKTMEVDVNLKQIPYHNRSLNNLSNALLMPTPYQTCKGDIULIAKVAAY 459
      402

384 YTGEDKFNV-----EAIKEDGSMFLFIDTPVNLMSDKNYFNPSKSNKIJVYRN 431
      384

460 YVNOQVVGIFKQDQNNQHQILAKVKKQDGTIAF-DTGALN-----NNSVYSLDKIVSNP 513
      460

432 E-----FYLRGK--ISDKGFWNEL-----RVNESVDVNNYLI--YGDLIHINT 470
      432

514 QNVLVSNFDLTSKQKQILKPEPAASVYSTKKTQILENIMNDLINOQKLVATVVD--NND 570
      514

471 RDEFNIKLVNKGQD--INDMGKKDYKANGFPD-----KVTMDGNGVYLTQGYSDLNAKA 521
      471

571 KEYIVANVNDQNNKIIPD-----SNDLPKGYIYHLAKVBNNDLNKVINLNDPELKQKI 623
      571

```


QY 522 VGVHGFYDYNKPEVNIDPKGNTSI EKADGKSVFVNINDK-----RNNGFDEIQRQH 575
 Db 624 IDKRDNLNLDLS--HPFTYDNDGNLEIH-----TQLANDLNDLDKQKALNNANKGIVVQ- 677
 QY 576 IYNGKEYTSFNDIQIIDKTLNITIKIVVDFAKNTTVE--FILKDCGSEVSELKPHVT 633
 Db 678 ---DIEH---EIVSDANGKVTIIPYNNLNNPPTKENTYTLCK----- 716
 QY 634 VTIQNGKEMSTIVSE-----DPIIPYKGELEKGYOPDGEIS----- 673
 Db 717 VTLKQNNQPNIDLIEBQSGDNHISFKKPTTAKTEN---DOYEISFSPNSLANKKIX 773
 QY 674 -----GPEKGDAGYVNLSDPTIKVPYKIEEKKEEENTPTDV 714
 Db 774 LTFKTDNNNTKTYEASIGLDGKA-----IFKTSDDAIFAPDHKYTLTFKIEADNKKVANI 829
 QY 725 SKKQNPQVNSHQLNESHKEDLQREHSHQKSDSKDVTATVLDKN----- 761
 Db 830 DEISPLDRIYVNGKGNVADN--KHFEKIPDQKNDLTAAYKDKNNNEIHVPIKTDKG 887
 QY 762 ---ISSKSTNNPNK 773
 Db 888 KTVVNPNNNLPDPNK 902

RESULT 11

US-10-282-122A-52048
 ; Sequence 52048, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zykend, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Foreyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See file wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: Patent version 3.1
 ; SEQ ID NO 52048
 ; LENGTH: 1191
 ; TYPE: PRT
 ; ORGANISM: Clostridium acetobutylicum

US-10-282-122A-52048

Query Match 5.1%; Score 207; DB 15; Length 1191;
 Best Local Similarity 19.6%; Pred. No. 0.0038;
 Matches 178; Conservative 139; Mismatches 275; Indels 314; Gaps 43;

QY 1 KIGTJASCKPKNLNGKKG-----SIKQDTGVHHHQNESI 39
 Db 120 RLKIDIOE-LFMDTGIGKEGYGIIIGCKIEAVLSGKPEERRALLBEAAGIVFKTKRVA- 177
 QY 40 KEKSFITIDRINSTRD-----FENKDKLKKKPREVD--FISEGK--RME 86
 Db 178 -EKLLENTQNLVRLNDILRTYERLEPLRIESEKAKFVELSDLTKEINTIIVSIDN 236
 QY 87 YDKYDDKGNITIA-----YDDGT-----DLAEYETKLEI--SKIYGLSP 126
 Db 237 IDYRLNDLKQKADLKLSDENVDKESISLEKATSELSDFDAKSSNKKTY-----E 292
 QY 127 SKDGFELIGKI--SNVSKAKVYGNKYISF-----IKATKYDFHSKTMTPDLY 175
 Db 293 SKSEHQILSEIELKEKTSNSDVAKNLYKEIEDLNSIVLTK-SRYEIQLTLETED-- 349
 QY 176 ANINDIYDGLAFAGDMRLFYDNDQKAEIKRME--KIKETSE-----YPPVSSGN 228
 Db 350 KNYNK-----ELSKINKSEKKNIDGLLEWEKSIKQYKNDIIDIISTISQNNN 400
 QY 229 VIELGEGDLSKKNKPNLTWMS--GKIVDSEKQOYLKNDILIRGYALKVTPYNG 284
 Db 401 EVVI-----LKKIEBESKLESIKRAGEGYKS-----LKINETKNTLSBELVKIN-- 448
 QY 285 KTMLENGVYSKEDIKIQANPNRLSETTYIADSRNVEDGSRSTOSVLMALDGNFI 344
 Db 449 --DKISGYENOIRERRSKISKLN--RIISDE----- 475
 QY 345 IRYQVFTFRKNDKGAIDKDGNLVDSKVLFGDKDEYTGEDKFNYEAIKEDGMLFI 404
 Db 476 ---EKLNEELNSKSKKLEANKMLTNL-----EKQYGYNK-SVKNLMOHYTKGVP 522
 QY 405 DTKPYNLSMDKNYFNPSKSKKIYVNPPEYLRGKISDKGFRWELRVESVD--NYLI 461
 Db 523 DYKPESSFVLBEVIRKVEFETA--BISGAISD-----ITLDDNIAKLIANTLK 573
 QY 462 YGDHIDNTRDNT--KLVKDG----- 482
 Db 574 SKNLGRATFLPLNITIKGRKLNISDTRHEKGTIGIASELIDYDSTFLPAVNYVIGRTYIV 633
 QY 483 DIMQGMKDYKANGFPDVTMDGNVYLQTYSDLMAKAVGHVFLVDNKPVEVNIDPK 542
 Db 634 DNMDSALKIANKISYFKIVTLTGEV-----VNG 663
 QY 543 GNTS--IEYADGKSVFVNINDKRNNGFDEIQRQHIIYNGKEYTSFNDIQIIDKTLNIX 600
 Db 664 GSLTGSGTYSKRAAGI-----GRKREIEELNLENNVSA--LEQSSNKIINX 710
 QY 601 IYVQDFAKNTTVEFILKDTGSEVSELKPHRYVTYIQNGKEMSTIVSEEDFILPYIGE 660
 Db 711 KVVKE-----LDNLCD-----LTDTI-HGEKIELTKIKERKSIDIESEK 750
 QY 661 LEKGQFDPGWEISGPEKGDAGYVNLSDPTIKVPYKIEEK--KEEENKPTDVSCK 717
 Db 751 LNKSYNTSVGBIG-----PIKEKINKHLEKLKVKSEENKA--LKL 789
 QY 718 KDNQVNSHQLNESHK-----EDLQREHSHQSD--STQDVTATVLDK 759
 Db 790 EAN--NNNLDELERKLDKNSKVLNNEIRIMSKVDKASDEMLMSTREIEKYKEM 846
 QY 760 NNISCK 765
 Db 847 HNMEK 852

RESULT 12
 US-10-744-616-7


```
/ Sequence 7, Application US/10744616
/ Publication NO. US20050026170A1
/ GENERAL INFORMATION:
/ APPLICANT: Patti, Joseph M.
/ APPLICANT: Foster, Timothy J.
/ APPLICANT: Hook, Magnus A.O.
/ APPLICANT: Eldhim, Delidre N.I.
/ APPLICANT: Perkins, Samuel L.
/ TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
/ FILE REFERENCE: P06283US2/BAS
/ CURRENT APPLICATION NUMBER: US/10/744,616
/ CURRENT FILING DATE: 2003-12-24
/ PRIOR APPLICATION NUMBER: 60/066,815
/ PRIOR FILING DATE: 1997-11-26
/ PRIOR APPLICATION NUMBER: 60/098,427
/ PRIOR FILING DATE: 1998-08-31
/ NUMBER OF SEQ ID NOS: 23
/ SOFTWARE: Patent Ver. 2.0
/ SEQ ID NO 7
/ LENGTH: 1166
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-10-744-616-7
```

```
Query Match          5.0%; Score 203; DB 17; Length 1166;
Best Local Similarity 21.6%; Pred. No. 0.0062;
Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;
```

```
QY 33 QNBSIKR-KSSFTIDRNIITRDPEKDLKLIKFR-----EVDFTS 78
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 226 KNNPELTKELVNDSTNDSTKRVATAPTVAKRVNAKPAVAPAAVASNNVDLIK 285
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 79 ETGKRMEEYDYKDDKGNIIAYDDGTDEYETE-----KLDEIKSIKYGVS 125
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 286 VTKQITK---VGDGKNVAAAHADGKQIEYDIEFTIDNKKYKGDITTYIDKNVIVSDLT 341
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 PSKD-----GHPEILKISNVSNAK---VYGNVNSYIEIKATKYDFHSK----- 168
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 342 DKNDPIDIDPGEVIAKGPBKATQIITYTFDYDKEDIKSRITLYSIDKTVVNE 401
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 ---TMTF---DLVANN-DIVDGLAFAGD---MELFYK-DNDQKAEIKRM-PEKIK 214
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 402 TSLNLFATAGKETSQNVTVDYODPVAH-GDSNIQSIPTKLEDEKQITIQIIVNLIKKS 460
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 215 ETKSEY---PYVSSYGVNIELEGEED--LSKNNKDLTKM-----ESGKIYSEKQ 261
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 461 AATNKVDIAGSOVDYDIGN-IKLGNSITIDQTEIKRVVNSDQOLPQSRITVDSQYED 519
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 262 YLLK-DNIIIRKGYALKVTTYPNGKTDMLGNGVSKEDIKIQKANPILALSETTIYA 320
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 520 VTSQFDN---KKSFSNNVATIDFG-----DINSAYI-----IKVSKYTPPS 558
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 321 DSR-NVEDGRSTQSVLMSALDGFNIIIRYOVFTFRKNDKG-----E 359
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 559 DEELDAQGTSMKTT--DKYGYNNVAGYNSFIYTSNDTGSGDGTVPBEKLYKIGYVWE 616
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 360 AIDKQGNLVTDSSKVLFGKDDKEYTGEDEKFNVEAIKEDGSMFLITDKPVNLSMDKNY-F 418
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 617 DVDKDD-----VQGTDSKEKMAVNL-VTLTYPDGT-----TKSFRDANSHYER 660
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 419 NPSKSNKIVV---RNPEFYLRGKIS-----DKGGSFNNELRVNBSVNNYLIYGLDHTD 468
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 661 GGLKQDEYTVKFEPTGYLPTFKVNGTTDGEKDSNSSVTVKINKG-----DMSJSD 712
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 469 -----NTDFNIIKLANVKG--DIMWGKDYRANGFPPKATVDMGQNVYLQIGYSPL 517
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 713 TSPYKPKYNLGDYVWEDNKGQIDANBPGIDVAVT-----LKDSGTGV-IGITTTTDA 766
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 518 NAKAVGVHQLYDVNVKPEVNIDPKNGTSIEY---ADGKSVNEN--INRKNNGPFGELQ 572
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 767 SGK---YKF-----TDLD-NGNYTVEFEETPAGPYPTVANKTTAADSDSGL----- 807
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 573 ECHIYINGKEYTSFNDIKQIIDKTLNIIKIVVDPAANT--TVKEPI---LNKDTGEVSEL 627
| : : : : : : : : : : : : : : : : : : : : : : : :
```

```
DB 808 -----TTTGVIKQADNMWTLD-----RGFYTKPKYSIGADYVWYDNNKDGKQDSTYR 851
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 628 KHR-VTYTONGK-EMSTIYSEDDFILPVKSGLEKYOQDGMIEISFEKKAQAGYI 685
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 852 KAIKQVTVTLQNEKEGVIGITTYIDN-----GK---YRPD----- 883
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 686 NLSKTPFIKPVFKTIEKKEENKPTFVSKKKQNPQV-----NHSQ--LINESHRKEDLQ 738
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 884 NDSGKY-KVIFKPKAGLQTYVTNTED-DKQADGGEVDTITDHDFTLDNGYFREDTS 941
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 739 REHSQKSDSTXD 751
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 942 DSDSDSDSDSDSD 954
| : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 13
US-10-032-585-7646
/ Sequence 7646, Application US/10032585
/ Publication NO. US20030180953A1
/ GENERAL INFORMATION:
```

```
/ APPLICANT: Terry, Roemer D.
/ APPLICANT: Bo, Jjiang
/ APPLICANT: Charles, Boone
/ APPLICANT: Howard, Busesey
/ TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
/ FILE REFERENCE: 10182-005-999
/ CURRENT APPLICATION NUMBER: US/10/032,585
/ CURRENT FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 8000
/ SOFTWARE: Patent In version 3.1
/ SEQ ID NO 7646
/ LENGTH: 1881
/ TYPE: PRT
/ ORGANISM: Candida albicans
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (1881)..(1881)
/ OTHER INFORMATION: X-any amino acid
US-10-032-585-7646
```

```
Query Match          5.0%; Score 203; DB 14; Length 1881;
```

```
Best Local Similarity 20.8%; Pred. No. 0.011;
Matches 172; Conservative 132; Mismatches 300; Indels 222; Gaps 39;
```

```
QY 22 KQDTGVV---HHQNEBSIKESSTFIIDN-ISTIRDPEKDLKLIKKGFEVDVF 76
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 1140 KATDEIEKTKHITDLOEHAQKQSQFSEHNDIKSNLDEANKELSD-NREKLSNLEKE 1198
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 77 TSETGRMEEDYKTDKGNIIAY--DDGTDEYETEKLDEIKSIKTYGVLSKQGHPEI 134
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 1199 KTELNNKLTQSEKISDLETVAISEDKSKIKHDEIDKREKIKLETTLKENEBEWFEX 1258
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 LCKISVNSKNAKRVYVGNVNSYIEIKATKYDFHSKTMTPDLVANNIDVGLAFAGDMRLF 194
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 1259 KEQLOVNDKCK-----ELEAC-----LKKLTKEKEINDLIRKLEAA----- 1297
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 195 VQDNOQKRAEIKRPEKIKETKSEYPVSSYGVNIELEGEGLSKNNKPNLTKMESGKIY 254
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 1298 KSDHDTERRKQSL-----LIEDTKSR-----SEKNVIKLE-----QIEKIK 1334
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 255 SDSEKQOYLLKNNIIRKGYALKVTTYPNGKTD-----LEGNVYSKEDIKQIK 305
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 1335 GERKE---VRD---IQSLAKTTDWEIKITTLKVLKESKDLERT--NKESVDTLK 1385
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 ANPNLR--ALSETTIYADSRNVEDGRSTQSVLMSALDGFNI-----IRYQVF- 350
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 1386 EVENLKEKISLEDDQKDDTTKXKEIAQLFETKSNLDSITWELEKTELELKKVNEELTE 1445
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 351 ---TFKNDKG-----EALDKQGNLVTDSSKVLFGKDDKEYTGEDEKFNVEAIKEDGSMFL 402
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 1446 ATSELTKLODNNQSLTEIEKTKAALTVSSK-----DLEVCQNGQKSELQ-----DSLK 1493
| : : : : : : : : : : : : : : : : : : : : : : : :
```

QY 403 FIDTPEVNLSDKATFNPBKSXKIYVRNPEFY-----LKGKISDKGGMFELRVNESVVD 457
Db 1494 SYKSLKPKENKYNKOTTSLSLKDSEIKKEIYVTLQTELDKRISEVEKERAMLSENSETV- 1552
QY 458 NYLIYGDHLIDNTRDPRNKLNVKQGDIMDMGKDYKANGFPDKVTMDGNVYLQGYSDL 517
Db 1553 -----IKESIS-----DKISLE-----SKY 1567
QY 518 NAKAVGVHYQFLYDVNKEPVNIDPKGNTSIBX-----ADGKSVFNNIDKNNPDEGEI 571
Db 1568 NS-----IKENHSKEITTHNEOKTSLKQDIAKLSQDHSAQOQLEDKENQKELKA 1618
QY 572 Q-EQHIIYNGKEYTSF-----NDIKQIID--KTLNKKIYVKDPAKNTTWKF-----ILNKD 620
Db 1619 SLEKFI-----NTESATISEEKNOQIKELSELTKSLKTELKTSGDALQSQSKYKTLKTKQSD 1675
QY 621 T-----GEVSELKPRHVVTIIONGKMSSTIVSEEDFIPVYKGELEKGYQGDWEISGF 675
Db 1676 TESKLEKQLEBE--KVASDLOTADEKLGKITERIAL-----KSELET----- 1717
QY 676 ECKKQAGVYINLSKDTFIKPFVKIIEKKEENKPTFDVSKKQDPQVNHQSOLNESHK- 734
Db 1718 --VKNSGLSTSELALFTKV--KSLKEKKEELQFLSGNKSKELEADYIQGHSDISEKGLXL 1774
QY 735 EDLQGEHNSQKSDTK-----DYATATYLDKNNISKSSTNNPNK 773
Db 1775 TDELKEKTKQFPDSSKKQLTELENDLST---KKELETEKTQTSKPK 1817

RESULT 14
US-09-742-096-3
; Sequence 3, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERGES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 200773US01V
; CURRENT APPLICATION NUMBER: US/09/742,096
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
; PRIOR FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: P. falciparum
US-09-742-096-3

Query Match 5.0%; Score 200.5; DB 9; Length 1786;
Best Local Similarity 20.5%; Pred. No. 0.015;
Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;
QY 6 AES--KFKN--LGNGKESGLKQDT-----TGVENHNOEN-----ESIK 40
Db 868 AESVTTFNSNILEIENITNTDITIEKLEELHENVLSALENTQSEBEKKEVIDIEEK 927
QY 41 EKSGSTI-----DRNISTRD--FEN-----KDLKKLIKKEKREYDDFTS 78
Db 928 EVATTLLETVQAEKESANTITEIFENLEENAVESNENVAENLEKLNMTVFNVLDAKVE 987
QY 79 ET-----GKRMF--EVDYK-----DDGK--NIIAYDGTDLKDET-----EKLD-- 114
Db 988 ETVESIGSLENNENDKAFSEIFPNVKGIOENLLT--GMFRSLETSLVIOSEKVDLN 1044
QY 115 -EIKSKIYGVSPSKDGHPEILGKISNVSKNAKV-----YGGNNYSIBIKATKYDFH 166
Db 1045 ENVVSIIDNINEMKEG---LANKLENISSTGCVQETVTEHVBQVNVYDVDPVPAKK----- 1097

QY 167 SKTMTFDDYANINDIVDGLAFAGDRLFLVKONDQKAKIKIRMPKIKETKSEYVSSY 226
Db 1098 -----DQFLGILNABAGL-----KEMFNNLEDVFKSESVDVTEVEIKDEPQKVEKERET 1146
QY 227 GNVIELGEGD-----LSKNKPNLTKW-----ESGKIYVDSQOQLLDNIIIRGVALK 277
Db 1147 VSIIIEEMENIVDVEEKEDLTDKMDAVEESIISSDSKSETESIKDK--EKDVSLV 1203
QY 278 VTYNPG-----KTDMLEGNVSKF--EDIAKIOKANPNL----- 310
Db 1204 VEEVDNDMDBSVEKVELKNNBEELMDAYEINITSKLIEFQELAVEADLKDMEK 1263
QY 311 -----PALSETTYVADSBNVEDR--STQSVLMSALDGFNIIRYOVFTPKANDKGEALDK 363
Db 1264 LKELEKALSE-----DSKEIIDIADKDTLEKYLEEEDH-----ITTLDEVELKDV 1309
QY 364 DGNLVTDSKVLVFKGDKEYTGGEKPFVVEALKEGSMFLDTKPVNL-----SM 413
Db 1310 BEDKIEKVSGL-----KDLBEDILKE--VKEIKELSSILEDYKELKTLETDLKEKKEI 1362
QY 414 DKATVPBKSXKIYVRNPEFYLRGKISDKGGMFELRVNESVVD-----NYLIYGD 465
Db 1363 EKDHFKEKEEABEIKDLADILKEVS-----SLEVEBEKLEBEVHEKVEHIIISGDA 1417
QY 466 HIDNTRDPRNKLNVKD--GDIM-----DMGKDYKANGFPDKVTMDGNVYLQGYSDLN 518
Db 1418 HIKGLEEDDLR--EVDLKGSIIDMLKGDMEIGDMDEKSLBEDVTYTLGSRV--SGLKQVL 1473
QY 519 AKAVGVHYQFLYDN-----VKPEVNIIDPKGNTSIEYADGKSVFNNPNK--RN 564
Db 1474 SSALGMDDEQKTRKKAQPKLEEVLLKEEVEEERKKIT---KKKVRPDIKKEPXD 1528
QY 565 NGFDEIOEHIYINGK-----EYTSFNDIKQIIDKTLN-----IKIYVKDPAF--N 609
Db 1529 EIVEVEMQEDIBEDVEEDIEDBEDKVEDIDEDIDEDIGDKQVADLIVQKREKLEK 1588
QY 610 TTVEKFIINKDTGE--VSELKPH-----RTVTIIONGKMSSTIVSEED 651
Db 1589 VAKKKKLEKVEEVSGLKGVDEVMKYVOKIDKXVDKVEKSKLESKNDVTNVKQND 1648
QY 652 FILPYKGBELKRGVQFDGMEISGFGKGDAGVNL-----SKDTFIKPFVK 698
Db 1649 FFSKV--KNFVKKYKFAAPFISAVAA--PASTVGFPTFSLSCVTTAASSTYILSKDK 1705
QY 699 KIEEKEEENKP-----TFDV-----SKKQDPQVNHQSOLNESHKEDLQK 739
Db 1706 TINKKE---RPFSFVFDIFPNLKHVYLQKKEKSKKNNVIEVT--NKAKKKGNQV 1760
QY 740 BEHSQKSDSTQDVATATYLDKNNISKSSTNNPNK 773
Db 1761 TNKTEKT-----TKVDKNNKVPKRRRTQSK 1786

RESULT 15
US-10-415-253-2
; Sequence 2, Application US/10415253
; Publication No. US20040067236A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Joe
; APPLICANT: Druilhe, Pierre
; TITLE OF INVENTION: Immunogenic Compositions Comprising
; FILE REFERENCE: B45250
; TITLE OF INVENTION: Liver Stage Malarial Antigens
; CURRENT APPLICATION NUMBER: US/10/415,253
; CURRENT FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: PCT/EP01/12349
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: EP00203724.0
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

LENGTH: 1787
TYPE: PRF
ORGANISM: K1 Parasite Clone
US-10-415-253-2

Query Match 5.0%; Score 200.5; DB 15; Length 1787;
Best Local Similarity 20.5%; Pred. No. 0.015;
Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

QY 740 EHSQSDSTKDVTATVLDKNNISSKSTNNPNK 773
DB 1761 TWKTEK-----TKVDKNNKVPKCRRTQSK 1786
Search completed: February 16, 2005, 19:32:38
Job time: 147 secs

QY 6 AES--KPKK-LGNGKESLAKDT-----TGVEHHQENE-----ESIK 40
DB 868 AESVTTFSNLEIQENTITNDTIEBKLELHENVLSALENTQSEKKEVIDVIEBK 927
QY 41 EKSFTI-----DRNISTRD--FEN-----KDLKKLKKKPREVDFTS 78
DB 928 EBAVATLLETVAEKSANTITEITFENLENAVESNENVAENLEKNETVFNVTLDKYE 987
QY 79 ET---GKME--EYDYKY-----DDKC--NIAVDDGTDLEVT-----EKLD-- 114
DB 988 ETVEISGSLENNEMDKAFSEIFDVVKGIQENLLT--GMFRSIFTSIVIQSEKVDLN 1044
QY 115 -EIKSKIYGVSPSKDGHFEIIGKISNVKNAKV-----YGNKYKSEIKATKYDFH 166
DB 1045 ENNVSSILDNINMKEG---LNLKLENISTEGVQETVTEHVEQNVYVDVDPAMK---- 1097
QY 167 SKTMTFDLYANINDIVDGLAFAGDMRLFYKNDQKAEIKIRNPEKIKETKSEYFVSSY 226
DB 1098 -----DPLGILNEAGL-----KEMFNLBDVFKSESDVITVEBIKDEBPQKEVERET 1146
QY 227 GNVIELGEGD---LSKNRPDLTYK-----ESGKIYDSEKQOYLKDNIIIRKGYALK 277
DB 1147 V8IIEEMENIVDLVEEKEDLTDKVIDAVESIEISDSKSTESIKDK--EKDVSIV 1203
QY 278 VTTYPG-----KTMLEGNVYSK--EDIAKIQKANPVL----- 310
DB 1204 VEEVDNDMDSEVEKYLEKNNEEBELMKDAVEINDITSKLIEGTOLNEVADLIDMKER 1263
QY 311 -----RALSETTYADSRNVEDOR--STOSVLSALDGNIRYQVTFKANDKGEAIDK 363
DB 1264 LKELEKALSE-----DSKEIIDAKDPTLEKIVEEHD-----ITTLDBEVVELKDV 1309
QY 364 DGNLVYDSSKLVLFCKDKEYTGEDKFNVBAIKEDGSMLEFIDTKPVNL-----SM 413
DB 1310 EEDKIEKVS DL---KDLIEDILKE---VKEIKELSEILBDYKELKTIETDILEEKKEI 1362
QY 414 DKQYFNPSSKSNKIYVANKPEFYLRGKISDGKGFNMLELVNESVVD-----NYLIYGD 465
DB 1363 EKDHFEKFEFEAEBEIKDLADILKEVS-----SLEVEEKQLEVEHLEKEVEHIIISGDA 1417
QY 466 HLDNTRDPNIXLVKND--GDIM-----DMGMDYKANGFPDKVTMDGNVYLQTVSDLN 518
DB 1418 HIKGLEEDDL- EVDLTKSILDMKGLDMLGDMKESLEBDVTTKIGERY--ESLKDYL 1473
QY 519 AKAVGVHYQFLYDN-----VKPEVNIDPKGNTSIEYADGKSVFVNINDK--RN 564
DB 1474 SSALGMDDEQMTKPKKAKQPKLEEVLLKEVEKEPKKIT-----KKQVRFDIKDKERPO 1528
QY 565 NGFDGEIQOHYIYNK-----EYTSNDIKQIIDKTLN-----IKIVKDPAR--N 609
DB 1529 EIVEVEMKDEIDIEEDVEBDEIDEDKVEDIDEDIDEDIDEDKDEVIDLIIVQEKRIEK 1588
QY 610 TTVKEFILNKDGE--VSELKPH-----RVTYIIONGKMSSTIIVSEB 651
DB 1589 VAKAKKKLREKVEGVSGLKGVDEVMKVVQKIDKEVDKESKALSKNDVTNVLKQND 1648
QY 652 FILPVYKSLKGYQDFGWEISGFEKQDAGYVNL-----SKDTFIKPVK 698
DB 1649 FFSKV--KNFVKKKVPAAPFISAVAA--FASVYVGFFTSLFSSCVTIASSYLISKVDK 1705
QY 699 KIEEKEEENK---TPDV-----SKKDNPOVNHSQLNESHKEDLOR 739
DB 1706 TINKKE---RPFYSFVFDI FKNLKHVLOOMKEKFSKEKNNNVEVT--NKAEKGNVQV 1760

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:09:54 ; Search time 181 Seconds
(without alignments)
2166.946 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGFIASKFKNLGNKGKGS.....ATVLDKNMISKSTNNPKK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: uniprot_03:*
2: uniprot_sprot:*
3: uniprot_trembl:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	2119	2 Q9AHT5	Q9AHT5 streptococc
2	4026	100.0	2140	2 Q97YR6	Q97YR6 streptococc
3	3793.5	94.2	2144	2 Q9S4M8	Q9S4M8 streptococc
4	3709.5	92.1	2144	2 Q8DQ7	Q8DQ7 streptococc
5	270.5	6.7	2649	2 Q7RAS7	Q7RAS7 streptococc
6	265.5	6.6	1642	2 Q81B84	Q81B84 plasmodium
7	264	6.6	3504	2 Q81L45	Q81L45 plasmodium
8	260	6.5	1850	2 Q7RGP2	Q7RGP2 plasmodium
9	259	6.4	2757	2 Q7RRR9	Q7RRR9 plasmodium
10	258.5	6.4	1811	2 Q7RRH9	Q7RRH9 plasmodium
11	257.5	6.4	2661	2 Q7RRS4	Q7RRS4 plasmodium
12	257	6.4	1389	2 Q7RRJ4	Q7RRJ4 plasmodium
13	255	6.3	1777	2 Q81JP4	Q81JP4 plasmodium
14	253.5	6.3	1127	2 Q9YVT6	Q9YVT6 melanoplus
15	252.5	6.3	2227	2 Q81I21	Q81I21 plasmodium
16	252.5	6.3	2273	2 Q7ROB6	Q7ROB6 plasmodium
17	251.5	6.2	2849	2 Q81HY4	Q81HY4 plasmodium
18	251.5	6.2	3381	2 Q81E64	Q81E64 plasmodium
19	251.5	6.2	3519	2 Q81E65	Q81E65 plasmodium
20	250	6.2	1474	2 Q81LU2	Q81LU2 plasmodium
21	250	6.2	1650	2 Q77J28	Q77J28 plasmodium
22	247.5	6.1	5767	2 Q81E25	Q81E25 plasmodium
23	246.5	6.1	2033	2 Q81M16	Q81M16 plasmodium
24	245	6.1	2269	2 Q81UA2	Q81UA2 plasmodium
25	242.5	6.0	2664	2 Q7REL0	Q7REL0 plasmodium
26	242.5	6.0	3317	2 Q8EWP8	Q8EWP8 mycoplasma
27	241.5	6.0	1033	2 Q81BB8	Q81BB8 plasmodium
28	241.5	6.0	2694	2 Q7RJ31	Q7RJ31 plasmodium
29	241.5	6.0	3063	2 Q6IMC1	Q6IMC1 plasmodium
30	240.5	6.0	2740	2 Q7RFS2	Q7RFS2 plasmodium
31	240.5	6.0	4433	2 Q81J15	Q81J15 plasmodium

32	240	6.0	1455	2 Q81K68	Q81K68 plasmodium
33	240	6.0	10061	2 Q81J21	Q81J21 plasmodium
34	239.5	5.9	2569	2 Q81B68	Q81B68 plasmodium
35	239	5.9	2586	2 Q7PDT7	Q7PDT7 plasmodium
36	238	5.9	1081	2 Q8XIL2	Q8XIL2 clostridium
37	238	5.9	3322	2 Q8XIL0	Q8XIL0 plasmodium
38	238	5.9	3628	2 Q968Y0	Q968Y0 plasmodium
39	238	5.9	3704	2 Q81KY8	Q81KY8 plasmodium
40	237.5	5.9	1104	2 Q7RS08	Q7RS08 plasmodium
41	237.5	5.9	3335	2 Q81C29	Q81C29 plasmodium
42	237	5.9	2511	2 Q81L44	Q81L44 plasmodium
43	237	5.9	5229	2 Q7RFP4	Q7RFP4 plasmodium
44	236.5	5.9	1114	2 Q97242	Q97242 plasmodium
45	236.5	5.9	1738	2 Q81AL5	Q81AL5 plasmodium

ALIGNMENTS

RESULT 1
Q9AHT5 PRELIMINARY; PRT; 2119 AA.
ID Q9AHT5
AC Q9AHT5
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Serine protease (Fragment).
GN Name:prta;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RX MEDLINE=21116976; PubMed=11179332;
DOI=10.1128/JAI.69.3.1593-1598.2001;
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
Langemann S., Johnson S., Koenig S.,
RT "use of a whole genome approach to identify vaccine molecules
affording protection against Streptococcus pneumoniae infection.",
RL Infect. Immun. 69:1593-1598(2001).
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
an amide bond (By similarity).
EMBL: AF291699; AKK19159.1; -.
DR HSSP; P00782; 2SER.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_SS3.
DR InterPro; IPR010259; ProC_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFA; TIGR01167; LPXG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN 1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN 1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN 1.
DR Cell wall; peptidoglycan-anchor; Protease.
KW NON_TER 1
FT

```

SQ SEQUENCE 2119 AA; 238227 MW; 51799B7F6B960A6A CRC64;
Query Match 100.0%; Score 4026; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 2e-161;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KGEIAESFKNLGNGKESLKKDTTGVHHHQQNEESIKESFTIDRNIISTIDPFENK 60
DB 1313 KGEIAESFKNLGNGKESLKKDTTGVHHHQQNEESIKESFTIDRNIISTIDPFENK 1372
QY 61 DLKKLIKKEKFEVDPTSETGKRMEEYDYKYDDKGNIIAYDGTDLLEYETELDKISKI 120
DB 1373 DLKKLIKKEKFEVDPTSETGKRMEEYDYKYDDKGNIIAYDGTDLLEYETELDKISKI 1432
QY 121 YGVLSPSKDGHEIIGKISNVSKNAKYVYGNKYKSIIEIKATKYDFHSKTMTPDLANIND 180
DB 1433 YGVLSPSKDGHEIIGKISNVSKNAKYVYGNKYKSIIEIKATKYDFHSKTMTPDLANIND 1492
QY 181 IYDGLAFAGDMRLFYKNDNQKKAETIKIRPEKIKETKSEYYPVSSYGVNIEIGEGDLSKN 240
DB 1493 IYDGLAFAGDMRLFYKNDNQKKAETIKIRPEKIKETKSEYYPVSSYGVNIEIGEGDLSKN 1552
QY 241 KPDNLTKNESGKIYSSEKQOYLKDNIIIRKGVALKTTYNPGKTMLEGNVYSKEDI 300
DB 1553 KPDNLTKNESGKIYSSEKQOYLKDNIIIRKGVALKTTYNPGKTMLEGNVYSKEDI 1612
QY 301 AKIQKANPRLALSETTIYADSRNVEDGRSTQSVLMSALDGNFIIRYQVTFPKANDKGBA 360
DB 1613 AKIQKANPRLALSETTIYADSRNVEDGRSTQSVLMSALDGNFIIRYQVTFPKANDKGBA 1672
QY 361 IDKDNGLVTDSSKLVLFQKDDKEYTGEDGFENVEAIKEDSSMLFIDTKPVLNLSMDKNYFNP 420
DB 1673 IDKDNGLVTDSSKLVLFQKDDKEYTGEDGFENVEAIKEDSSMLFIDTKPVLNLSMDKNYFNP 1732
QY 421 SKSNKIYYANPEFYLRGKISDCKGFENWEIARVNESVVDNLIYGDLLHIDTRDFNLIKLVNK 480
DB 1733 SKSNKIYYANPEFYLRGKISDCKGFENWEIARVNESVVDNLIYGDLLHIDTRDFNLIKLVNK 1792
QY 481 DGDINDMGNKDYKANGFPDKVTDMDNVYLTQGYSDLNKAVGVHYQFLYDNVKEEVNID 540
DB 1793 DGDINDMGNKDYKANGFPDKVTDMDNVYLTQGYSDLNKAVGVHYQFLYDNVKEEVNID 1852
QY 541 PKGNISIEYADKSVFENINDKRNNGFDEIOBHIIYINGKEYTSNDIKOIIDKTLNLIK 600
DB 1853 PKGNISIEYADKSVFENINDKRNNGFDEIOBHIIYINGKEYTSNDIKOIIDKTLNLIK 1912
QY 601 IYVKDFARTTYKEFLINKDTGEVSELKPHRYVTITIQNGKEMSSITVSEEDFILPYKGB 660
DB 1913 IYVKDFARTTYKEFLINKDTGEVSELKPHRYVTITIQNGKEMSSITVSEEDFILPYKGB 1972
QY 661 LEKGIQFDGWEISGFEKGDAGVINALSKDTFIKPVFKKIEEKEEENKPTFDVSKKON 720
DB 1973 LEKGIQFDGWEISGFEKGDAGVINALSKDTFIKPVFKKIEEKEEENKPTFDVSKKON 2032
QY 721 POWNSQLNESHKEDLOREESHQSSTKDYATATLDKNNISSTKTTNPNK 773
DB 2033 POWNSQLNESHKEDLOREESHQSSTKDYATATLDKNNISSTKTTNPNK 2085

RESULT 2
097RY6 PRELIMINARY; PRT; 2140 AA.
AC 097RY6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Serine protease, subtilase family.
GN OrderedLocustNames=SP0641;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=1313;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Knouri H.M., Wolf A.M., Ullrichack T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang P., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 295:498-506(2001).
CC -! SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (By similarity).
DB EMBL, AB007373; AAK74791.1; -.
DB PIR, F95074; F95074.
DB HSSP, P00782; 2SRT.
DB MEROPS, S08.064; -.
DB TIGR, SP0641; -.
DB GO, GO:0009986; C:cell surface; IEA.
DB GO, GO:0005618; C:cell wall; IEA.
DB GO, GO:0008233; F:peptidase activity; IEA.
DB GO, GO:0004283; F:subtilase activity; IEA.
DB GO, GO:0006508; F:proteolysis and peptidolysis; IEA.
DB InterPro, IPR010435; DUF1034.
DB InterPro, IPR001899; Gram_pos_anchor.
DB InterPro, IPR003137; PA.
DB InterPro, IPR000209; Pept_S8_S53.
DB InterPro, IPR010259; Pept_inh_S8A.
DB InterPro, IPR001680; WD40.
DB Pfam, PF06280; DUF1034; 1.
DB Pfam, PF00746; Gram_pos_anchor; 1.
DB Pfam, PF02225; PA; 1.
DB Pfam, PF00082; Peptidase_S8; 1.
DB Pfam, PF05922; Subtilisin_N; 1.
DB PRINTS, PR00723; SUBTILISIN.
DB TIGRFAMs, TIGR01167; LPTYG_anchor; 1.
DB PROSITE, PS00847; GRAM_POS_ANCHORING; 1.
DB PROSITE, PS00137; SUBTILASE_HIS; UNKNOWN 1.
DB PROSITE, PS00138; SUBTILASE_SER; UNKNOWN 1.
DB PROSITE, PS00678; WD_REPEATS_1; UNKNOWN 1.
KW Cell wall, Complete proteome, Peptidoglycan-anchor, Protease.
SQ SEQUENCE 2140 AA; 240426 MW; FA44ADB2938B334 CRC64;

Query Match 100.0%; Score 4026; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2e-161;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KGEIAESFKNLGNGKESLKKDTTGVHHHQQNEESIKESFTIDRNIISTIDPFENK 60
DB 1313 KGEIAESFKNLGNGKESLKKDTTGVHHHQQNEESIKESFTIDRNIISTIDPFENK 1393
QY 61 DLKKLIKKEKFEVDPTSETGKRMEEYDYKYDDKGNIIAYDGTDLLEYETELDKISKI 120
DB 1394 DLKKLIKKEKFEVDPTSETGKRMEEYDYKYDDKGNIIAYDGTDLLEYETELDKISKI 1453
QY 121 YGVLSPSKDGHEIIGKISNVSKNAKYVYGNKYKSIIEIKATKYDFHSKTMTPDLANIND 180
DB 1514 YGVLSPSKDGHEIIGKISNVSKNAKYVYGNKYKSIIEIKATKYDFHSKTMTPDLANIND 1573
QY 241 KPDNLTKNESGKIYSSEKQOYLKDNIIIRKGVALKTTYNPGKTMLEGNVYSKEDI 300
DB 1574 KPDNLTKNESGKIYSSEKQOYLKDNIIIRKGVALKTTYNPGKTMLEGNVYSKEDI 1633
QY 301 AKIQKANPRLALSETTIYADSRNVEDGRSTQSVLMSALDGNFIIRYQVTFPKANDKGBA 360
DB 1634 AKIQKANPRLALSETTIYADSRNVEDGRSTQSVLMSALDGNFIIRYQVTFPKANDKGBA 1693

```

QY 361 IDKDNLTVDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMLEFIDTKPVNLSMDKNYFNP 420
 DB 1694 IDKDNLTVDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMLEFIDTKPVNLSMDKNYFNP 1753
 QY 421 SKSNKIYVNPPEFYLRGKISDYGKGFNWEI RVNESVVDNYLIYGDLDHIDTRDFNKLANK 480
 DB 1754 SKSNKIYVNPPEFYLRGKISDYGKGFNWEI RVNESVVDNYLIYGDLDHIDTRDFNKLANK 1813
 QY 481 DGDINDMGKDYKANGFPDKVTDMGDNVYLQYGSIDLNAKAVGHYQFLYDNVKEPVND 540
 DB 1814 DGDINDMGKDYKANGFPDKVTDMGDNVYLQYGSIDLNAKAVGHYQFLYDNVKEPVND 1873
 QY 541 PKGNTSIEYADGKSVFENINDRNNNGFDEIOEHIYNGKEYTSFNDIKQIIDKTLNIX 600
 DB 1874 PKGNTSIEYADGKSVFENINDRNNNGFDEIOEHIYNGKEYTSFNDIKQIIDKTLNIX 1933
 QY 601 IYVKDPARTTYKEFLINKDTGEVSELKPHRTVVTIIONGKEMSTTVSEDFILPYKGE 660
 DB 1934 IYVKDPARTTYKEFLINKDTGEVSELKPHRTVVTIIONGKEMSTTVSEDFILPYKGE 1993
 QY 661 LEKGVOFGDMEISGEGKKDAGVYNLSKDTFIKPVFKKIEEKEENKPTDVSCKKN 720
 DB 1994 LEKGVOFGDMEISGEGKKDAGVYNLSKDTFIKPVFKKIEEKEENKPTDVSCKKN 2053
 QY 721 PQVNSQLNESHREKEDLOREHSHQSDSTKDYATATVLDKNNISSTKSTNNPNK 773
 DB 2054 PQVNSQLNESHREKEDLOREHSHQSDSTKDYATATVLDKNNISSTKSTNNPNK 2106

RESULT 3

Q9S4M8 PRELIMINARY; PRT; 2144 AA.

ID Q9S4M8
 AC Q9S4M8:
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
 DE Cell wall-associated serine proteinase precursor PrtA.
 GN Name:PrtA;
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_Taxid=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3.B:
 RX MEDLINE=21585565; PubMed=11728722;
 RA Bethe G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,
 Zydek G.;
 RT "The cell wall-associated serine protease PrtA: a highly conserved
 virulence factor of Streptococcus pneumoniae.";
 RL FEMS Microbiol. Lett. 205:99-104(2001).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 an amide bond (By similarity).
 DR EMBL; AF127143; AAD48399.1; -.
 DR HSSP; P00782; 2SBR.
 DR MEROPS; S08.064; -.
 DR GO; GO:0009986; C:cell surface; IEA.
 DR GO; GO:0005618; C:cell wall; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; F:protease activity and peptidolysis; IEA.
 DR InterPro; IPR010435; DUF1034.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Pept_S8_S53.
 DR InterPro; IPR010259; Procl_inh_S8A.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF006280; DUF1034; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; peptidase_S8; 1.
 DR Pfam; PF05922; Subtilisin_N; 1.

DR PRINTS; PR00723; SUBTILISIN.
 DR TIGRAME; TIGR01167; LPXTG anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN 1.
 DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN 1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN 1.
 DR Cell wall; Peptidoglycan-anchor; Signal.
 FT SIGNAL 1
 FT CHAIN 20 2144
 FT PrtA
 FT cell wall-associated serine proteinase

SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;

Query Match 94.2%; Score 3793.5; DB 2; Length 2144;
 Best Local Similarity 94.7%; Pred. No. 1.3e-151;
 Matches 732; Conservative 16; Mismatches 24; Indels 1; Gaps 1;

QY 1 KLGELAESKFKNLGKSGSLKQDTTGYEHHHQENBESIKESSTTIDRNISTIDFENK 60
 DB 1339 KLGELAESKFKNLGKSGSLKQDTTGYEHHHQENBESIKESSTTIDRNISTIDFENK 1398
 QY 61 DLKXLIKKKFRVDDFTSETKRMEYDYKYDDKGNIIAYDGTDLLEYETKLDDEIKSI 120
 DB 1399 DLKXLIKKKFRVDDFTSETKRMEYDYKYDDKGNIIAYDGTDLLEYETKLDDEIKSI 1457
 QY 121 YGVLSPSKDGHEFLIGKISNYSKNAKYVYGNYSIKIKATKYDPHAKTMTPDLYANIND 180
 DB 1456 YGVLSPSKDGHEFLIGKISNYSKNAKYVYGNYSIKIKATKYDPHAKTMTPDLYANIND 1517
 QY 181 IYDGLAFAGDMFLFYDNDQKAEIKINPEKIKETKSEYFYVSSYGNVIELEGEDLSKN 240
 DB 1518 IYDGLAFAGDMFLFYDNDQKAEIKINPEKIKETKSEYFYVSSYGNVIELEGEDLSKN 1577
 QY 241 KEDNLTNRESGKIYSDSEKQOYLKDNIIIRGVALKATTYAPGTMDLEGNGVSKEDI 300
 DB 1578 KEDNLTNRESGKIYSDSEKQOYLKDNIIIRGVALKATTYAPGTMDLEGNGVSKEDI 1637
 QY 301 AKIQKAPNLRALSETTIYVADSRNVEDGRSTQSVLMSALDGFNIIIRYQYFTRANDKGA 360
 DB 1638 AKIQKAPNLRALSETTIYVADSRNVEDGRSTQSVLMSALDGFNIIIRYQYFTRANDKGA 1697
 QY 361 IDKDNLTVDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMLEFIDTKPVNLSMDKNYFNP 420
 DB 1698 IDKDNLTVDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMLEFIDTKPVNLSMDKNYFNP 1757
 QY 421 SKSNKIYVNPPEFYLRGKISDYGKGFNWEI RVNESVVDNYLIYGDLDHIDTRDFNKLANK 480
 DB 1758 SKSNKIYVNPPEFYLRGKISDYGKGFNWEI RVNESVVDNYLIYGDLDHIDTRDFNKLANK 1817
 QY 481 DGDINDMGKDYKANGFPDKVTDMGDNVYLQYGSIDLNAKAVGHYQFLYDNVKEPVND 540
 DB 1818 DGDINDMGKDYKANGFPDKVTDMGDNVYLQYGSIDLNAKAVGHYQFLYDNVKEPVND 1877
 QY 541 PKGNTSIEYADGKSVFENINDRNNNGFDEIOEHIYNGKEYTSFNDIKQIIDKTLNIX 600
 DB 1878 PKGNTSIEYADGKSVFENINDRNNNGFDEIOEHIYNGKEYTSFNDIKQIIDKTLNIX 1937
 QY 601 IYVKDPARTTYKEFLINKDTGEVSELKPHRTVVTIIONGKEMSTTVSEDFILPYKGE 660
 DB 1938 IYVKDPARTTYKEFLINKDTGEVSELKPHRTVVTIIONGKEMSTTVSEDFILPYKGE 1997
 QY 661 LEKGVOFGDMEISGEGKKDAGVYNLSKDTFIKPVFKKIEEKEENKPTDVSCKKN 720
 DB 1998 LEKGVOFGDMEISGEGKKDAGVYNLSKDTFIKPVFKKIEEKEENKPTDVSCKKN 2057
 QY 721 PQVNSQLNESHREKEDLOREHSHQSDSTKDYATATVLDKNNISSTKSTNNPNK 773
 DB 2058 PQVNSQLNESHREKEDLOREHSHQSDSTKDYATATVLDKNNISSTKSTNNPNK 2110

RESULT 4

Q8DQP7 PRELIMINARY; PRT; 2144 AA.
 ID Q8DQP7
 AC Q8DQP7;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DT Cell wall-associated serine protease PrtA (EC 3.4.21.-).
 GN Name=prtA; OrderedLocustName=aprt0561;
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NC NCB1_TaxID=171101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;
 RX DOI=10.1128/JB.183.19.5709-5717.2001;
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA Dehoff B.S., Estrem S.T., Filtz L., Fu D.-J., Fuller W., Geringer C.,
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
 RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA McAhren S.M., McHenry M., McMaster K., Mundy C.W., Nicas T.I.,
 RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
 RA Sun P.-M., Winkler M.B., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Roatock P.R. Jr., Skatrud P.L.,
 RA Glas J.I.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
 RL J. Bacteriol. 183:5709-5717(2001).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 an amide bond (By similarity).
 CC EMBL; AB008434; AAK9365.1; -.
 CC PIR; A97942; A97942.
 DR HSPD; P00782; 2SPT.
 DR MEMOPS; S08_064; -.
 DR GO; GO:0008986; C:cell surface; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilisin activity; IEA.
 DR GO; GO:000508; P:protease activity and peptidolysis; IEA.
 DR InterPro; IPR010435; DUF1034.
 DR InterPro; IPR010499; Gram_pos_anchor.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Pept_S8_S53.
 DR InterPro; IPR010259; Prot_inh_S8A.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF06280; DUF1034; 1.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; peptidase_S8; 1.
 DR Pfam; PF05922; Subtilisin_N; 1.
 DR PRINTS; PRO0723; SUBTILISIN.
 DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
 DR Cell wall; Complete proteome; Peptidoglycan-anchor.
 KW KW
 SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBCS03A0C CRC64;
 Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
 Best Local Similarity 92.4%; Pred. No. 4,5e-148;
 Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

QY 241 KPDLITKMSGKIYSDSEKQYLLKDNILKRGVALKVTYNNPGKTDMLKNGVYSKEDI 300
 DB KPDLITKMSGKIYSDSEKQYLLKDNILKRGVALKVTYNNPGKTDMLKNGVYSKEDI 1637
 QY 301 AKIQKAPNLRALSTTTIYADSRNVEDGRSQTQVMSALDGNITIRYOVFFKMDKDEA 360
 DB AKIQKAPNLRALSTTTIYADSRNVEDGRSQTQVMSALDGNITIRYOVFFKMDKDEA 1697
 QY 361 IDKGNLVTDSKVLFGKDKXEYTGEDKFVNEAIKEDGSMFLDTKEVNLSDMKRYFNP 420
 DB IDKGNLVTDSKVLFGKDKXEYTGEDKFVNEAIKEDGSMFLDTKEVNLSDMKRYFNP 1757
 QY 421 SKSNKIYVRNPEFYLRGKISDGKGFNWEIYNVESVVDNYLLYGDIAHINTRDENIKLVK 480
 DB SKSNKIYVRNPEFYLRGKISDGKGFNWEIYNVESVVDNYLLYGDIAHINTRDENIKLVK 1817
 QY 481 DGDIDMKMDKYKANGFPDXYTDMGNGNYLTQVSDLNAAKGVHYQFLYDQVKEVND 540
 DB DGDIDMKMDKYKANGFPDXYTDMGNGNYLTQVSDLNAAKGVHYQFLYDQVKEVND 1877
 QY 541 PKGNTSIEYADGKSVFVFNINDKRNNGFGEIOEHIYINGKEYTSFNDIKQIIDKTLNIX 600
 DB PKGNTSIEYADGKSVFVFNINDKRNNGFGEIOEHIYINGKEYTSFNDIKQIIDKTLNIX 1937
 QY 601 IVKDFANNTYKEFLNKDITGEVSELKPHRVTVTIQNGKENSITVSEEDFILPVYKGE 660
 DB IVKDFANNTYKEFLNKDITGEVSELKPHRVTVTIQNGKENSITVSEEDFILPVYKGE 1997
 QY 661 LEKGYPFGWETSGEGKKDAGVYVNLNISKDFIKRVFKIEKKEENKPFVDSKKDN 720
 DB LEKGYPFGWETSGEGKKDAGVYVNLNISKDFIKRVFKIEKKEENKPFVDSKKDN 2057
 QY 721 PQVNHSQLNESHKEDLQREHSHQKSDSTKVATVLDKNNISSKSTNNPK 773
 DB PQVNHSQLNESHKEDLQREHSHQKSDSTKVATVLDKNNISSKSTNNPK 2110
 RESULT 5
 QY 07RAS7 PRELIMINARY; PRT; 2649 AA.
 AC 07RAS7;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein.
 GN Name=PY06422;
 OS Plasmodium yoelii yoelii.
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NC NCB1_TaxID=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=1236865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angiuoli S.V., Suh B.B., Koell T.W., Pertea M.,
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 RA Shallow N.J., van Aken S.E., Riedmiller S.B., Feldblyum T.V.,
 RA Cho J.K., Quekkehush T., Sedegah M., Shoat A., Cummings L.M.,
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria
 parasite Plasmodium yoelii yoelii.";
 RL Nature 419:512-519(2002).
 CC -1- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
 preliminary data.
 CC EMBL; AABL01002172; EAA18637.1; -.
 DR InterPro; IPR011591; Botulinum.
 DR ProDom; PD001963; Botulinum; 2.

KM Hypothetical protein.
SQ SEQUENCE 2649 AA; 309810 MW; E7207F344643AC24 CRC64;
Query Match 6.7%; Score 270.5; DB 2; Length 2649;
Best Local Similarity 20.6%; Pred. No. 0.0053;
Matches 206; Conservative 160; Mismatches 327; Indels 307; Gaps 49;
QY 1 KLGEIAESKPKNLGN-----GKEGSLKDKDTGVGHHQHEESIKESSTPIIRNISTIRD 56
DB 1044 KNNMLIERRYRDNMLHQVKA KNNKKDLSNLYNSKSGDNYVQRRSITTEPFTTIS 1103
QY 57 FE-----NKDLKLKKKPREVVDFTSETGKMEEDYKDDKGNIIAYD----- 102
DB 1104 FEDKIKILQLQNEBRLKKIEKLYDEKEKYNKY-KMEKIKKXQD---NLFEATDGHIEK 1159
QY 103 -GTDLEVEREKDELKSKYGVLSPKDGHFELIGKISVSKAKAYYGN--NYSIEIK 159
DB 1160 LHCELENNSKONELLKINL-----KDKDMKILEESQICNN-NVSHSEHLNHTLVDVK 1211
QY 160 ATKYDFHSKTMFEDLYANINDIYDGLAFAGDM-----RIFVKDNDQKAE-IKIRMEPI 213
DB 1212 NSENNTKHTNNNDVSLSDSITDKKKIKQIVLQDQFLMKLKEIKQMEPLSKNVELN 1271
QY 214 KETKSEYFVSSYGAVIELGEGDLSKKNPDL----- 245
DB 1272 KLLNMKKNHINERYERSIDKLEKSIDKLEQNLKLEQICDLKEKNIMEKAAQLRDESS 1331
QY 246 ---TMESEG-----KIYSD----- 256
DB 1332 NTTTISIDSGTTINNEIKIMKEIEALYDKIKLKSNEBKTNNKINILNTLTKTSNEOS 1391
QY 257 ---SEKQOYLKDNILIRKGYALKVTYTPGKTMLEGNVYSKEDIKIQKAPMLRAL 313
DB 1392 IELANKIKLLENKMLQKNEKSI-----NDLKKNIKYD-EDILKENSITKPSIT- 1442
QY 314 SETTIYADSRNVEDGRTQSIVLSALD---GENIIR--YQVTFKRNKDG-----EAID 362
DB 1443 ---TFPLANKEKIEHNTVKQTDDESTQIDSDYDMRIKGYDISNHKONSECVQDQITEID 1500
QY 363 KD-----GNL---VTDSKVLFGKDKKEYT--GEDKFNVAIKEDDSMLFIDTKPV- 409
DB 1501 KNNYEKCTPNLKEQTQDIDTNIIVLISVDKKEMTKKNQDDYN---ILEDNNKIKNETNSMF 1557
QY 410 NLSMDKNYFNPSSKNKIYVRNPEFY-----LRGKISDGKG---FMWELRVNESVVD 457
DB 1558 NNRKENITNMLKNEKKYVANNIFETDISRANLQMFSAANGNESFPNVEVANN--H 1614
QY 458 NYL-----IYGDLDHIDNTRDFNIK---LNVKXGDJMDGMDKYKANGPDDKVTMDGNY 508
DB 1615 NYTNMKIKINAYSSHLSN---NISEBHYVQNTESLQ-----NVNKKYVNIKYAPIDEV 1664
QY 509 YLQTSIDMLAKAVGCH-----YQFLDYKPEVNIIDP 541
DB 1665 YNRN-----NLQFISLKHQDENNRRTISENGRTSYNSVEKDNDSNRYNNSNKEVYIEN 1719
QY 542 KNGTSIEVADGKSVFENDK-----RANGPDEGIOEHIYINGKEYTSFNDIKQIIDK 595
DB 1720 NNNNSIKNSD-----NVNKKYIECLANKINEFESANNESTLANTFENNSSTNDLKIIIE 1773
QY 596 TLNIKI---VVKDPANITVKEFILN-KDTGEVSELKPRVTVITIQNGKEMSSTIVS-EE 650
DB 1774 DNNTIANNKTIENY-NDQDLKAYVLSQKTNNNNEHKPND-NMTNEKKKKGEAMITIDIKN 1831
QY 651 DFLIPYKGELEKGYCFDGEWISGFGKKDAGYVNLKDTFIKVPFKIIEKKEBENKP 710
DB 1832 NEVLPLYTK--IEKCVLSD-----EKBAAGTKKNNK-----KSCQCKSKNSKNS 1872
QY 711 ---TFDVSCK-----KKDNPQV-N-H-----SOLNEHR 723
DB 1873 VNNMRRTNIVYRPSBSISAKTEBPANKKAPNLMHKTNGKLSLVNDIKLYANKI 1932
QY 734 KEDLQREHSQKSDSTKDVY---ATVLDKNNISSKSTYNN 770

DB 1933 KEELKKNIS-----KQITNLEITTTIKKKNPISKSLIANN 1966
Query Match 6.6%; Score 265.5; DB 2; Length 1642;
Best Local Similarity 20.4%; Pred. No. 0.0048;
Matches 190; Conservative 133; Mismatches 313; Indels 297; Gaps 46;
QY 4 EIAESKPKNLGKSGSLKDKDTGVGHHQHEESI---KESSSTPIIRNISTIRDLENK 60
DB 104 EIKRRNNRRNNRRNNNSNSNFINDKEKKNIRITMLKHKRKYPDPTNIS-----DEK 158
QY 61 DLKDLIKKKFREVDFTSETGKRM-----EEDYDYDDKGNIIAYDDGTDLEETEKLDE 115
DB 159 NIKR-----KKNIDINKYKKGKMLIPSYEEDKYSKNGSEFNID-----DE 201
QY 116 IKSICYGVLSPSKDGHFELIGKISVSKAKAYYGNKYKSIKATKYDFHSKTMTPDLY 175
DB 202 M-----DKNSNYNISGK-----NY-STENNNDKNNFHIRSTL-EKQ 235
QY 176 ANINDIYDGLAFAGDMRLFKVNDQKKAIEIKIRMEPIKETSEYFVSSYGAVIELGSG 235
DB 236 DKINN-----EKKEY-----STKSIDLRRESISHSGKE-----NNNIDKLKCH 275
QY 236 DLSKRPDLTMESEGKIYSDSEKQOYLKDNILIRKGYA----- 275
DB 276 NLIRNKNMLVNPDRNK-----RKSNFIFLHNMYKKKGNGKNTNINSVDIENIHDEKDK 330
QY 276 ---LKVITYNPGK---TDMLEGN-----GVYSKE-DIAKIQKANPMLRAL 314
DB 331 NEILANNLDINSYNEBKCSYNNMKENDFPVIGISKQPNVDSERKYNKYEKOHNIYNDS 390
QY 315 ETTIYADSRNVEDG-RTQSIVLSALDGNITIRYQVTFKRNK----- 357
DB 391 DDGKSIDRTNYYDNNSSSHVSLSLV--FNESNFMLEDEKIKIDKEVVLIEKSLERNNS 448
QY 358 -----GEALDKD--GNLVTD-----SSKLVLFGKDKKEYTGEDKFNVAIKE 397
DB 449 EYIENELKEGETHKDELEKLVLDMTIKD INEKKKVVILYHKOYS-----NYEKSSIS 503
QY 398 DGSMLFIDTKPVNLSDMKRYFNPSKSNKIYVRNPEFYLRGKISDGKFMWELRVNESVVD 457
DB 504 EN-----IDIGPTNRSGLVDGNNTKQMTIHNEENDR-----DKRNDN-----INN--D 544
QY 458 NYLIGDLDHIDNTRDFNIKLVNKGQDINDWKKDYKANGFPKVPKVTMDGNY-----YL 510
DB 545 NIIINNDDNNINNNNNVNVK--ESNEDIF-----YKANGLSKINDIVONIRIKIKKIS 596
QY 511 QNGYSDMLAKAVGCHVQFLYDNY-KPEVNIIDPKG---TSIEVADGKSVFENIDKRN 565
DB 597 NSGNVNIIDAN-----LSDGISNDEANLKVNTSMYTKLYYDDK-----GENTON 642
QY 566 GPDGEIOEHIYINGKEYTSFNDIKQIIDKTLINIKIVVDKPARNTTVKEFILNKDGEVS 625

```

Db      643 TF-----KSAAYDKNKKLLIENDYKLMHGNNN-----QKEMIRND--EKYLLINGINNSPK 691
Qy      626 ELKPRHVTYTIQ-----NGKEMSSITVSEEDFILPVYKGLKGYQFDGWEISGFEGKCD- 680
Db      602 DLNDRNSNIDIQVNNNNNNNNNTTFYDAWEYSIGNDKEQYNNQVGD-HVINGVSNKKN 750
Qy      681 -----AGYVINSKDTFTI-----KVPF-- 697
Db      751 MQTNCSSNNKKNYINNDRWNHKLKXYFDWMKPGVVDNIINIKENELVLOKNEKKDIFNK 810
Qy      698 -----KRIEKKKEENK--PTEDVSKKKNPQVNH-----SQNESH---RKEDLQRE 740
Db      811 SHGNEKIEPLKDPNNKNVAVTILKKNVDDIHEDVKSVINIGNEIDENKIKKKNLINS 870
Qy      741 EHSOKSDSTKD-VTAVTLDKKNNISKSJTNNPN 772
Db      871 INGEKDKGNKNIINTSYSKNLLSKSYAKEPH 903

```

RESULT 7

```

ID      081145 PRELIMINARY; PRT; 3504 AA.
AC      081145:
DT      01-MAR-2003 (Tremblrel. 23, Created)
DT      01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE      Hypothetical protein.
GN      ORFNames=PL14_0404;
OS      Plasmodium falciparum (isolate 3D7).
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_taxid=36329;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22555705; PubMed=12368864; DOI=10.1038/nature01097;
RA      Carlton J.M., Hall N., Fung E., White O., Bertman M., Hyman R.W.,
RA      Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA      Eissen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA      Chan M.S., Nene V., Shalimov S.J., Suh B., Peterson J., Angiuoli S.,
RA      Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA      Martin D.M., Fairclamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA      McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA      Venter J.C., Carucci D.O., Hoffman S.L., Newbold C., Davis R.W.,
RA      Frazer C.M., Barrett B.,
RT      "Genome sequence of the human malaria parasite Plasmodium
RT      falciparum."
RL      Nature 419:498-511 (2002).
DR      EMBL; AE014822; AAN37017.1; -.
DR      HSP; Q931C2; IMX.
KW      Hypothetical protein.
SQ      SEQUENCE 3504 AA; 408308 MW; B8454D855BE4F0 CRC64;

```

Query Match 6.6%; Score 264; DB 2; Length 3504;

Best Local Similarity 21.2%; Pred. No. 0.014; Mismatches 346; Indels 252; Gaps 43;

Matches 196; Conservative 131; Mismatches 346; Indels 252; Gaps 43;

```

Qy      11 KNLGNGKSGSLKCKDTTGVHHQEN--EESIKK-----SSFTIDRNISITRDEPKDL 62
Db      1432 KNSHNSSENMITEYKGNKSPQENIDNIIPEKNGKNNQSGNSDNHIMTEKNNKNS 1491
Qy      63 KKLKKKREVDVDFSETGKMEYDYKDK---GNIIAVDGTDLLETETKLDKDEIKS 118
Db      1492 QQ-----NSDHIIMTEYEKNNQSENTDNDIMTEGYKNNKSEKKEEDIAS 1539
Qy      119 KIYGV-----LSPSKD-GHEILIGKISNYSKNAKYVG---NNKSIETIKTKDF--- 165
Db      1540 --YELDKRISHENQEHFTYPSRINKEFTNIVYNIVSGNEKGLSVKNISDIIIPD 1597
Qy      166 -HSKMTFEDLVANINDIVDGLAFAGDMRLFVNDQKAEIKIRIPEKIKETKSE---YP 221
Db      1598 GKGNKIQDIIIEEN-----GKNNKFEINIEDKISDXTQ-KSKISHNAGHFTP 1647
Qy      222 YVS-----SYGNVIELGEGDLSKNRPDNLTKMESGIY-----SDSEKQYLL 264

```

```

Db      1648 YESGNKISDENDEVYNISDINTNDOE--EVESKRIFETNDNINKHISSDNNKINOK 1705
Qy      265 KONILIRKGYALKVTVYNGPKTD-MLGNGVYSKEDIKQKAPNLR-----ALSETTI 318
Db      1706 QNNIINESQDKNIIVNH--KLDKILKHEHTSDESLKIKENKNTKKNKSINNEKI 1763
Qy      319 YADSRNV-----EDGRSTQVLSALDGFNIIRYQVETFRON----- 355
Db      1764 EEEKENVKNDETITGKKEENESDCLKIOKISNENINKNILYTDVYNDKSYNAOGTGH 1823
Qy      356 -----DKGEALD---KDNLY----- 368
Db      1824 ENDETNGTNSINDGLDNVAKIDQYISKGENILOENKEDILPSVTINNSLGDKYKENIS 1883
Qy      369 -TDSKLYLFGKDDKEYTGEDEKFNV-----EAIKEDSGMLFIDTK-----PVNLS- 412
Db      1884 PEDIKMEVAHKNIQNTISEDELGTQKDNERNKEDKSPNGVEENHQNEDKIGEVNLSN 1943
Qy      413 MDKNYFNPSKSNKIYVNRPEFYLRGK-ISDKGFMWELR--VNESVVDNY-----L 460
Db      1944 MNMESNIGNSDTI--NOHLNKGNIHHKGNVNSETNMTNNGTQNIISNEQPEKNI 2000
Qy      461 IYGDHIDNTDPRNFKLVKNGCDIDMGKMKQYKANGFPDKTDMGNGVYLLQGYSDLAK 520
Db      2001 IRGD-DIDKKNMENVKIDETGN-----NIKINKYDNKAKILNELIIRKQGTDDSDAD 2052
Qy      521 AVGVHYQFLYDNVK--PEVINDPKNTSIEYDVGKSVNINDKRNNGFGEIOEHYIYN 579
Db      2053 DISTGSDKMQOIEKKNENIKHKNVNV-----KDKI--SNDKENIYPPF-HKHDITIS 2105
Qy      580 GKRYTSFNDIKQIIDKTINIKIYVDFPANTYVKEFILNKDTGEVSELPKPRVTVIIONG 639
Db      2106 DNKKKEFNNVLEIPKGNH-----LDKKEITFQVEEKSIGQDKSMENNVSF--NDG 2157
Qy      640 KEMSTIYSEDFILPVYKGLKGYQFDGWEISGFEGKQAGYINLSK----- 689
Db      2158 KDIIH---IQEBDI-----KKNINNVNDKHSKSKNNLHIDEPKRYEKEIKKH 2204
Qy      690 ---DPIKVPFKLIEKKEENKPT-----PVSKKKNPQVNHQSLNESHKEDLQRE 741
Db      2205 EIAHDIDIKERIEQDENSNNKPESENILVDVNAQDDK-----NISKLTDLHDOE 2256
Qy      742 HSQKSDSTKDYATVTLDKNNISKS 766
Db      2257 KGTNNDSVVE-----HNVS DKT 2273

```

RESULT 8

```

ID      07RG2 PRELIMINARY; PRT; 1850 AA.
AC      07RG2:
DT      01-MAR-2004 (Tremblrel. 26, Created)
DT      01-MAR-2004 (Tremblrel. 26, Last sequence update)
DE      Hypothetical protein.
GN      Name=PY04304;
OS      Plasmodium yoelii yoelii.
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_taxid=73239;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=17XNL;
RC      PubMed=12368865; DOI=10.1038/nature01097;
RA      Carlton J.M., Angiuoli S.V., Suh B.B., Koali T.W., Pereira M.,
RA      Silva J.C., Ermolaeva W.D., Allen J.B., Selengut J.D., Koo H.L.,
RA      Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA      Shalimov S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA      Cho J.K., Quackenbush J., Sedegah M., Shoabdi A., Cummings L.M.,
RA      Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA      Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA      van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA      Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,

```

RA Carucci D.J.
 RT "Genome sequence and comparative analysis of the model rodent malaria
 parasite *Plasmodium yoelii* yoelii.",
 RL Nature 419:512-519 (2002).
 CC -!- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 preliminary data.
 CC EMBL: ABL01001295; BAA16146.1; -.
 DR Hypothetical protein.
 KW Hypothetical protein.
 SQ SEQUENCE 1850 AA; 215313 MW; 00366A8D5CE21629 CRC64;

Query Match 6.5%; Score 260; DB 2; Length 1850;
 Best Local Similarity 19.2%; Pred. No. 0.0094;
 Matches 198; Conservative 154; Mismatches 379; Indels 298; Gaps 44;

```

QY 6 ASKPKNLGNG-----KEGSLKDDTGVHHHGHQENESIEKESFTIDNISTIR----- 55
DB 62 SBEKNDQKGIYKFNSTNKKSTNDSSYDNYLADYINE-----DNNLNSFNSVCIN 115
QY 56 ---DFEN-KOLKKLIKKF--REVDDFTSETGRMEYDYK-----YDDKNIITAYD---D 102
DB 116 SKLDYNNIIRKSLKGNFNNNNNDDESDSDYDLYGFKNIRIRSKESFPARDKID 175
QY 103 GTDLEYETEKLDEIKSKIYGVLSF-----SKDGHPI----- 134
DB 176 NAEKYSLEKNTIDN--IGPIAPINSISGDNIKKGIKTKTCTNISEYNSMNNKIEVYATS 233
QY 135 ---LCKISVSKAKAYYGNNTKSIKATKIDFHSKMTF----- 172
DB 234 LNNVNTKENDKYNMSEYKNDYKINTKQSEINQREKNIYKSPSTAILSRKSDNDN 293
QY 173 ---DLVANIINDYDGLAFAGDM-RLFYKNDQKAAIK----- 206
DB 294 LCDNDMDIYANMTDLAKMTMGGDMKGLYTQDGNINKTPKTSKSYFELKREYAKGN 353
QY 207 ---IMPEKIKETKSEY-----PYVSSY-----GNVIELEGDSLKNKPNLTMSGKTY 254
DB 354 ISELTEPLMLKDKEDYSGNTPTIAYLGGQPNKSLNGEDVLENKISFNK-ELDKOY 412
QY 255 SPSSEKQYLLKKNIIIRKG-----YALKVTTYNGKTM--EGNGVYSKDIK 302
DB 413 SKENFEPKPFINTKEDPEPEAEETIEENFNIIKONSFSKRNHSHSNNTID 472
QY 303 IOKANEN-----LRALSETT-----IYASDRVED-----GRSTQSVLMSAL 339
DB 473 VAGQANPNDKDELDIRRVSVTSKATNFKTEETVDMKSRPIKLSNKSYSRKSTNNA 532
QY 340 DGFNIIRYGVTFK-----MNDKGEALDKDGNLVTDSSKLVLFKDKDEYTGEDKEN 391
DB 533 NGDLILE-KVNTFKINNNNDVSTFKIFSCNNFNIMTTTFKONLF--KEPKYSKLDHNE 589
QY 392 VEAIKEDGSMLEF-----DTKPVNLSMDKYNFPKSKKIYRANPEFYIRKGISDK 442
DB 590 INFEKYGGMLEYYISINDKGINENENIGLSKNITFFNTVEDNDQIDRIDTIPLNRLSM 649
QY 443 GGFNMLRVNESV-----DNVLYLGGDLHTDNTRDFNKLNVKQDLM--DMGKDKYKXG 496
DB 650 YGKSYDIDINEGINNNNDKXNINNPNTLGSKF-----IPSSVAINOMQKGDINN 704
QY 497 FPDKVTMDGNVYLQGYSDNLNAKAVGHYQPLVNVKPEVNDIPKGNSTIEVADKSVV 556
DB 705 IPIINRKSNTIIGNSNSDIYKNTNNINY-----NNKFNNSNDPFGSGVQIVDTGENIN 759
QY 557 FN-----INDKANG-----PDGIRQEHYIYNGEYI 564
DB 760 FRNSSYESKIRIMEKYINDKNNNAIRKSNYINKMTTHDNTFGGDT-----INKANTN 813
QY 585 SFN-DIKOIIDKLNTKIY-----VKDFARTVVEF-----ILNKDTGEVSELK 628
DB 814 NFHIIDPNLLINNIRKINSYKGNINKNVEYEBONTSEQPKNDITNYSSGNGCD-N-872
QY 629 PHRVTVTIQNGKEMSSITVSEEDFILPVYK--GELEKGYQFDGWEISGF--EGKDA-- 661
  
```

```

DB 873 NSKVINLMEKSNFTNFKSIISENDGFIPLNKYIANRLSETADP-GYKESPYMESKSDENH 931
QY 682 -----GYVINLSKD---TFIKPVFKKIBKBEKENKPTDPVSKK 717
DB 932 KNFGHTLGNNDINNENNNNGNHDITKKYVDFPKKSFDDNI-DVNEIKNKILNDYAKY 990
QY 718 KNPQVNSQLN-----ESHKREDLQREHSHQKSDSTKYATATV-----D 758
DB 991 NNNNSENKANNMMEKKNSTFYKDPNEDTLAKYIHQIKSGITVEVEXILSHKEINFEINGE 1050
QY 759 KNNISKST 767
DB 1051 KNNIILKKT 1059
  
```

RESULT 9

```

QY 07RRR9 ID 07RRR9 PRELIMINARY; PRT; 2757 AA.
AC 07RRR9;
DT 01-MAR-2004 (TEMBLrel. 26, Created)
DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Rhoptry protein.
GN Name=PY00649;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Anguilo S.V., Suh B.B., Kooij T.W., Pierce M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Sengen J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., Van Aken S.E., Riedmuller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoab A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sindén R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Jansz C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.
RT "Genome sequence and comparative analysis of the model rodent malaria
  parasite Plasmodium yoelii yoelii.",
  RL Nature 419:512-519 (2002).
  CC -!- CAUTION: The sequence shown here is derived from an
  EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
  preliminary data.
  CC EMBL: ABL0100178; BAA17609.1; -.
  DR InterPro; IPR001623; DnaU N.
  DR InterPro; IPR006499; ReticulocytBP.
  DR TIGRfam; TIGR01612; 235da-fam.1.
  DR PROSITE; PS00636; DnaU_1; UNKNOWN_1.
  SQ SEQUENCE 2757 AA; 324532 MW; 55AD73DE2B9EFA37 CRC64;

Query Match 6.4%; Score 259; DB 2; Length 2757;
Best Local Similarity 20.6%; Pred. No. 0.017;
Matches 198; Conservative 163; Mismatches 326; Indels 276; Gaps 43;

QY 18 EGSLLKDDTGVHH-----HQENESIEKESF-----TIDNRI----- 51
DB 542 EAGVKKSYSVKMKRSILINKKLEENNEKYIKETQIKDLPFNOYKLTIDENIYINKLK 601
QY 52 -----STIRDFENK-----DKKLIRKKFREVDDFTSETGRMEY----- 87
DB 602 LEIKERIKNISDNEIYKKAVDKTIENNAIYIDELATSPYQVEHLKSIDTIYSTIK 661
QY 88 -----DYKYDKNIIAYDGTDEYETE-----XUDEIKSKIYGVLSPSKDG 130
DB 662 SELSQIYEDDIDKLTVNELSIYQENDIDNVEDKTKLDLQSKIDNYSKIQNMETATVS 721
QY 131 HFEIIL-----GKISNVSNAKATV-YGNNTKSIKATKYFHSKMTTFDLVANIINDVGL 185
DB 722 HLTNITETNKKLSDTIVEIKYIYGEISK--DLNKLTEDFPKNKK--ELSNKIND----- 772
  
```

```

QY 186 AFAQDMRLFVKQNDQ-----KKAELIKIRMPKTI-----KETXSEYVSSYGVNIEL 232
Db 773 -----YAKENDQIANVYSKSISEIRHNYNDQISIDNTKEBEAKQNTDQSKETYSITIS 824
QY 233 GEGDLSKN-----KPDNLTKME-----SGKISYDSEKQOYLKQNIILRKQALAV 278
Db 825 KEDETSKITINEVKNMKDEFLSKYDKYKINEFNCKENVDSEHTQFELTKIKAEVSEDTL 884
QY 279 TTY-----NEGKTDMLEGNVYSKE--DIAKIOKANPNLRALSET-----TIYAD 321
Db 885 SIYXKFNKDSKLINMETNNSIEKQYNTITLAKVDYIVCSSTESIKGNHKNQITILKE 944
QY 322 --SRVEDGRSTQSVLMSALDGFNIIRYQVTFPKXNDKEAIDKQNL--VTDSKVLV 376
Db 945 KLNQNIKITIKETNTEISYKDKFE-----NTLJINKINIELDKAF--KQASLNDYSSNNELMQ 999
QY 377 PEKDKKEYTGEDEKFNVEALKEQDSMLFID-----TKVYNSMK--NYFNYSKN--KIYVKN 430
Db 1000 YFNNLKNANQKKNEN-----MAYHOFDEKEKAVNNIIOKIEDINKNINIEITAYT 1050
QY 431 PEYFLNGKISDKGCFNMEL-----RVNESVVDNYLIYGDLDHIDTRDPNLIKLVKQDG 483
Db 1051 SIYNSIEEIEENIEKNIESTLNTNVELEKVTNTNPFKIKELKLYNFSDFG-----KEGN 1105
QY 484 IMDWCKDYKANGFPDKVTMDGNVYLTQYSDLNAAKAVGHYQFLYDNVKEVNIIDPK- 542
Db 1106 I-----KYTNVVKIKODIKVNOQIDHINIELSEIKKSENVVDKMAQINLTLEKY 1157
QY 543 -----GNTSIEYADG--SVFNINDKRY-----NCFDEIOEQHIYNGKETSINDK 590
Db 1158 ADTALSNVDSEIGIEKKQNIIVTKIDKKNIYEINLSEISKIE---KOQTSLEKY 1212
QY 591 QI-----IDKTLN 598
Db 1213 GINLSYGVQNLGTLFLEQIDEEKKSEHTIKAMEAVMEDLDNKKKSQEIENEMGIEMDN 1272
QY 599 IKIIVKDFARNTTVKPEFLNKDTGE--VSELKPHRVTVTTIONGKMSSTIVSEEDFILPVY 657
Db 1273 KEMEVLNISHDDDKYHTISKHNENISDIRKSLKI-IDQ-----FSRESIDINDI- 1322
QY 658 KGELEKGY---QEDGMEISGFEQKXQDAGVINLSKDTFIKPVFKIEEKEENKPTPV 714
Db 1323 KEELQNVSESQKHSEINQYLSK--IANIYNILKMLKIKIINKKYEYNTIEK----- 1375
QY 715 SKKKDNPQVYNSQLNESHREKEDLOREHESQKSDST---KDVATATVLD---KNNTSSKST 767
Db 1376 NKNKINDELSENEKLIKIKIKODISLKECKSKISTEITLDDKDIDGCIQNIKVLKTHISEET 1435
QY 768 TNN 770
Db 1436 NTN 1438

```

```

RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldlyum T.V.,
RA Cho J.K., Quackebush J., Sedegah M., Snodhi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser L.W., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.,
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RT Nature 419:512-519 (2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBD whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001591; EAA17051.1; -.
KW Hypothetical protein.
FT NON TER
SQ
SEQUENCE 1811 AA; 210756 MW; E76A90F7392BCA0A CRC64;

Query Match 6.4%; Score 258.5; DB 2; Length 1811;
Best Local Similarity 21.7%; Pred. No. 0.011;
Matches 179; Conservative 117; Mismatches 312; Indels 215; Gaps 40;

QY 11 KVLGKGGSLKQDTTGVHHQENES-----IKKSSFTIDNISTIRDFEKND 61
Db 555 KNINNSQHSNIEVNNRHHIDNNNDKNTVNSNNHILKNDETKIDINVDNANTYTGFE 614
QY 62 LKKLTKKRFREYDFTSETGKMEBYDYKDDKGNIIAYDDGTLEY-----ETEKLE- 115
Db 615 ILS-DKEDFKLEFLTNKMGSKFLPEKMLDCSKQWQIDLQREYINNNIELKLE 673
QY 116 -----IKSKTYVLSPSKDGHEFLIGKISNVSKN-AKYVGNVYSIEIKATKYPH 166
Db 674 KKKYLEKCIKNNKIKELMCGRYVGLNV--KICNMKKKIANEFLNNIINKNHEVTLTYE 731
QY 167 SKTWTF-DLYANINIDVQGLAFAGDMRLFVNDONDKAEIKRIMPEKIKETSEYVSS 225
Db 732 RSAMCTQDLTA-----QLYVIGLANK--KLEMKIKINEFK----- 764
QY 226 YGNVIELEBGDSLKNKPNLTGMESEK-----IYDSEKQOYLKNDITLKGVAL- 276
Db 765 --NNID-GVANSQENESDAKKKOSKKGKGYVHSLIYGFNKKSH--KQN--KKYGSIN 815
QY 277 ---KYTTNPGKTMLEBNGVYSKEDIKQKANNLALSTTTIYADSRVDEGRSTOS 333
Db 816 NDTKQTT-EPSSKD-----NLKKKENERITANNNEDEKIKEN 853
QY 334 VLMSALDGFNIIRYQVTFEK--WNDK-----GEAIDKQNLVYTSKVLFGKD--KEYT 385
Db 854 KQKG-----NNTENTVNTKNIENQYKXKNSDEDTKRSNIRTDKANNRPFQKNSYNEKT 908
QY 386 GEDKFNVEAIKEDGMLFIDTKPVNLSDMKYFNPS-KSNKIYVNPPEF---YLKGIS 440
Db 909 VEDK-----EIKETKNSVNEYYSNNSYENNSYLDNNSFKNNLKNKSEIV 953
QY 441 DKGGFNNELRVNESVVDN--YLIYGDLDHIDTRDPNITKLVYKDGIMQMGMDYKANGP 498
Db 954 EKSNNSN---IKKKYVDNGGSGFFGSF-----FMFK-----SDYKKG-- 987
QY 499 DKVTDMDGNVYLTQYSLNKAAGVHYQFLDYNKPEVNIIDPKGNTSIEVADGSVFN 558
Db 988 SQIESNNSAIYVPSDSNSN---GSDYSFKYNNKEDKQKQKNSKKSNSAFAFK-IFN 1043
QY 559 INDKRNGPDEIQ---EQHIYNGKETSFPNDIKQIIDKTLNIKIIVKDFARNTTVKEF 615
Db 1044 LSRKKKKKKQKQKPSLSKSEDKIKKKDKNTDNGSVNENDDK-----YNNESSNQ 1091
QY 616 IANKDTGEVSELKPHRVTVTTIONG-----KEMSITIYSEEDFILPVYKGELEKYQFPG 669
Db 1092 SSNEEDGHINNYKKN-----NGNKKKANNKSGNGLNNKSSSLEBEKFEDEEFQI-- 1142
QY 670 WEISFEQKQAGYINLSKDTFIKPVFKIEEKEENKPTPVDSKKKDNQVHNSQLN 729
Db 1143 -----LKKEVGHHTIEDVEKSIIFERFDEDE-----FDIGTKNSSSCSNS--N 1187

```

QY 730 ESHRKKEDLOREHESOKSDSTKVAT-VLDKNNISKSTNNP 771
 DB 1188 DSNDSD-----SNKSIHQSDGTGTQITKSSSKSEKKP 1224

RESULT 11

Q7RMS4 PRELIMINARY; PRT; 2661 AA.
 ID 07RMS4
 AC 07RMS4
 DT 01-MAR-2004 (TREMBLrel. 26, Created)
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Rhocty protein (Fragment).
 GN Name=PY02104;
 OS Plasmodium yoelii yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OK NCBI_TaxId=73239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=1236865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angluoli S.V., Suh B.B., Kooij T.W., Petrea M.,
 Silva J.C., Ermolaeva M.D., Allen J.E., Sengenut J.D., Koo H.L.,
 Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
 Cho J.K., Quackenbush J., Sedegah M., Shoabli A., Cummings L.M.,
 Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 Cunningham D.A., Preiser P.R., Bergman L.W., Valdiva A.B.,
 van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 Carucci D.J.;
 RA "Genome sequence and comparative analysis of the model rodent malaria
 parasite Plasmodium yoelii yoelii.";
 RT Nature 419:512-519(2002).
 RU -1- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL: AABL01000575; EAA21526.1; -
 DR InterPro: IPR011561; Pox_11_rel.
 DR InterPro: IPR006499; ReticulocyteBP.
 DR ProDom: PD014111; Pox_11_rel; 1.
 DR TIGRFAMs: TIGR01612; 235Rda-fam; 1.
 FT NON TER 2661 2661
 SQ SEQUENCE 2661 AA; 312269 MW; 0BD02AD0B712253B CRC64;

Query Match 6.4%; Score 257.5; DB 2; Length 2661;
 Best Local Similarity 20.0%; Pred. No. 0.018;
 Matches 179; Conservative 151; Mismatches 294; Indels 271; Gaps 41;

QY 22 KKDTGVHHQNEESIK-EKSSFTIDNIS-----TRDPENK-- 60
 DB 427 KKKLETTITLANKNEETVLEKEIRLEFKYISDEVAEKHVEELKIKLKEITIKYKKE 486
 QY 61 -----DLKKLTKKREVDFTSETSGKMEBYDYKTD-----KGNIIAYD 101
 DB 487 YIKKAIIDKKALENNKKYIDELGKNTPPQIIEYVKKDITITIKSELSEIKGNIVELY 546
 QY 102 DGTDLFEYFEKDEYSK-----IYGVLSPPSGHFEITIGKISNYSKN----- 144
 DB 547 NELSSVVOGENTIDPIPKKELETKSEINDVYNNKICNMKIEVEELTKNIEYSKNELSNL 606
 QY 145 -----AKVYVG--NNYSIKIKATKVDGHSKWTMPDIYANINDIVDSLAFAGDMRLPVKN 198
 DB 607 IIEIITIKTGEIDN-----DLNLTKEFKYKKEQ--ELSKIND-----TYKEN 647
 QY 199 DG-----KKAETIKITMPERI-----KETSSEYFVSSYGVNLELGGDLSK----- 239
 DB 648 DQLSVYQSITIEIRKHYNDQINVDNTKGAEAKONTQOSKEVAKKIPITKSGSKFINEVT 707
 QY 240 -KKPNULTGME-----SGKIYDSKCOOTLKDNII-----LRKGYALKVTTYNGP 284
 DB 708 AMKNELRLTKKYIDPDGAYKSKANSHEKITELVKKIRIISKYLNNYERK--FNNS 764

QY 285 KTDMLEGVNYSKE--DIAIKOKANPNRALSETTIVADSRVVEDRGSTQVLAALD-G 341
 DB 765 KLLIITETKSIBEEKKNINTLAKVDYIKVCYHT-----NELISFNKQITLKKLDKN 819
 QY 342 FNIIIRY-----QVTFK---MNDGEALDKGNIV-----TDSKVLV-----G 378
 DB 820 INTVKENNSIDLKLYNNKENILTDKTELDKFTDVALANDHESNNNELMKYFNNLKANG 879
 QY 379 KDDK-----EYTGEDFNNEATIKEDGSMLEFIDKPNVLSMDKNYFSPSKNKIYVNPERY 434
 DB 880 KKKKMLYQQFGEKKALIDDIKKNADINENYSKEVALIYASIVNISDEMERIEKSIES 939
 QY 435 LRGKISDKGFMWELRVNESVVDNYLIYGDLIHDITRPFNILQANVKGDIIMDKMKDYKA 494
 DB 940 INTQYIER-----VKTNTVNLNEIKGKLKQYOPHFGKEENK----- 977
 QY 495 NGFPDKVTDMGDNVYLQGYSDLNKAVGVHYQFLYDNVKEPVNIDPKGNTSIEYADGS 554
 DB 978 --YPDEINKIKNEI-----KTVG---QQIDQHNKLEIKKNSGSXNDIEIKGT 1021
 QY 555 VVF-NINDKRNNGFGEIQEOHIYNGKEYTSFN-DIQIIDKTINIKIVDPAINTV 612
 DB 1022 DKFENVTDK-----TLYNKPPEIEEKIKNI-VTKIDEKKNYI 1059
 QY 613 KEFINKDTGVESELKPHRVT-----VTIQGKEMSSITVSEEDFILPVYKGELEKGY 666
 DB 1060 K--INKLDEISEIEKNNTSLKKYKDINVSQGLGFLFQID----- 1101
 QY 667 FDGWEISGEGCKDAGVYINLSKDTFIKPV-FKKIEKKEEENKPTD---VSKKDN 720
 DB 1102 -----EEKKGAGHNIK-SMEEYMDLDNKKKSGOIEENMKINMDIKYINKEMBA 1151
 QY 721 PQVHNSQLESIRKEDLOREHESOKSDSTKVATVLDKN---ISSSTNNP 772
 DB 1152 INISHDDQYH--NISKGEKISD-----HKNSLIKITIQEFTESNN 1194

RESULT 12

Q7RPS4 PRELIMINARY; PRT; 1389 AA.
 ID 07RPS4
 AC 07RPS4
 DT 01-MAR-2004 (TREMBLrel. 26, Created)
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Hypothetical protein.
 GN Name=PY01465;
 OS Plasmodium yoelii yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OK NCBI_TaxId=73239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=1236865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angluoli S.V., Suh B.B., Kooij T.W., Petrea M.,
 Silva J.C., Ermolaeva M.D., Allen J.E., Sengenut J.D., Koo H.L.,
 Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
 Cho J.K., Quackenbush J., Sedegah M., Shoabli A., Cummings L.M.,
 Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 Cunningham D.A., Preiser P.R., Bergman L.W., Valdiva A.B.,
 van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 Carucci D.J.;
 RA "Genome sequence and comparative analysis of the model rodent malaria
 parasite Plasmodium yoelii yoelii.";
 RT Nature 419:512-519(2002).
 RU -1- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL: AABL01000391; EAA20802.1; -
 DR InterPro: IPR011591; Botulinum.
 DR ProDom: PD001963; Botulinum; 1.

KM Hypothetical protein.
 SEQ SEQUENCE 1389 AA; 164816 MW; 21851CA630F5D9F CRC64;
 Query Match 6.4%; Score 257; DB 2; Length 1389;
 Best Local Similarity 20.4%; Pred. No. 0.0089;
 Matches 193; Conservative 139; Mismatches 288; Indels 326; Gaps 48;

17 KEGSLKQDTGVVHHQNEESIKESFTIRNISTIRD 56
 354 KKEAQCKEKSSSEHHEKNI FYEKEILEASKYDVSYFENKEIKKDKSGSNYNKGVIPSS 413
 57 FENKDLKKLKKKFRVDDFTSETGRMEBYDYKDDKKNIIAYDGDGLVETEKLDRI 116
 414 EINKKQKQKINKKCNKVFSSHSDT---EKTEY---SNITS-----DLYNQKDKK 457
 117 KSKIVGLSPSKDGHFEILKISNYSKNAKAYVGN-NYSIKIKAT-----KY 163
 458 KLNITVGL---SKEDY---SNKQKQDEVEYVYDNDKTDIIOVNDKVVLDISPETHKY 510
 164 DHSK-TMTFDLYANINDIVDGLAFAGDMRLVVKNDQKKAIR--MPEKIKETSE 219
 511 KVRSNIIITINKIKNIKIEKI-----EINLITVQLKIPGRILENVILPHK----- 556
 220 YVYSSYGVNI-----ELGEGDLSKNKPDNLTQMSGKIYDSEK-----QQYL 263
 557 ---YSNDILKIQIKTISEENENDIKTNKE---TKNATSYQQSDKNSQPSDNKV 607
 264 LKDNIIILKRGVALKV-----TTYNPKGTDMEGNYVSKED- 299
 608 IKKMIDRTGSIKCYVYLPLNNINHEILNRLLINSTYE-----MYEKENL 657
 300 ---IAKIOKANPNIALSETTIYADSRNVEDGRSTQVL-MSALDGRN--ITRYQ-V 349
 658 LPDITVYLQ--SKSISSEIKETMYISIKCVANGIYVLPILNRNVINDFENKMNINSQNN 715
 350 FTFK-----NMDKEAIDKO-GNIVTDSKLVLFCKDKKEYTGEDKFNVEAIKEDGSM 402
 716 STFKPYYVSLYND---IDREIGNLI-MTKLI-----NKIKIKKLPKPKGYI 759
 403 ---FIDTPVNLISMDKN-----YFNPSKSNKIY 427
 760 INDICGRFFDKNNVLCVNDNRKNSGNGINQIGNLEIDHIGHETANYIFILMSNFII 819
 428 VAMP-----EFYIRKGISDKGKGNMELRVNESVVDNYLIYGLDHLDN 469
 820 VSSSLDLPESHLPPIBQVYVDFIEKMKELKLFIMFLAITETACIMLCFL----- 872
 470 TDFPNIKLNVKGDIMDMGKDYKANGFPDKVLTDMGQVYVLTQVSDLAKAAGVHYQFL 529
 873 ---FINNAQI---DSLFT---KNLKAVALQKENDQNNFFFTLYTISYE-RVQNNKHCPIY 923
 530 YDNVAP-----EVNIDPKGNTSIEYAD--GKSVENINDKRNNGFGEIOEQHIYI 578
 924 LNVSPFLIPRFQSEHNLN-----NISHADVEGHNLNGTSDBRHRIKKEEIIKKKKE 976
 579 NKEKTSFNDIKQIIDKTLINIKIVKDFKRNNTVKEFIANKDTGEVSEKLPKRVVTION 638
 977 NG-----IK-LKKTINDNDFIYDVARQSI-FLYN-----AIKNDALAKNIEN 1018
 639 GKEMSTIVSEDF-----ILPYKGELEK---GYQF-----DCM 670
 1019 NYDKKKQIITYTSYTKQFQDYIINILNRYNOLLQKEKKEYSFRREVETLOSNIRDOQ 1078
 671 EISGEKKQDAG--YVINTSKDTFIKPVFKKIEKKEE---ENKPTDVSKKKQNPQVN 724
 1079 DVKSYFGKNDNNIIFSGGLNQ-----FDKCKEKKRBEIGIKSSNNKEMQDNKYTDNYQNS 1131
 725 HQALNESHAKEDLQREHESQSDSTYQVATVLDKKNISKSTTN 770
 1132 KQTFSDKPKFQNIYTDENIQLN-----NNFSNTQVANN 1165

Q813P4
 ID Q813P4 PRELIMINARY; FRT; 1777 AA.
 AC Q813P4;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein PFE1095w.
 GN Name=PFE1095w;
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_Taxid=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
 Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
 Budge C.O., Burrows C., Cherevach I., Chillingworth C.,
 Chillingworth T., Christodoulou Z., Clark L., Clark R., Corron C.,
 Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
 Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
 Harper D., Hauser H., Hornby T., Holroyd S., Horrocks P.,
 Humphray S., Jagsels K., James K.D., Johnson D., Kerhornou A.,
 Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
 Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
 Oliver K., Ormond D., Price C., Quail M.A., Rabinowitz E.,
 Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
 Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
 Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
 Raulston J.E., Craig A., Newbold C., Barrell B.G;
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
 RL Nature 419:527-531(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
 Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AL929353; CAD51583.1; --
 DR InterPro; IPR011591; Botulinum;
 DR ProDom; PD001963; Botulinum; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 1777 AA; 213320 MW; 244467CFE190C522 CRC64;

Query Match 6.3%; Score 255; DB 2; Length 1777;
 Best Local Similarity 20.7%; Pred. No. 0.015;
 Matches 186; Conservative 152; Mismatches 324; Indels 226; Gaps 44;

4 EIAESKFNKLGNGKGSFKDQTVGVHHQNE--ESIKESFTIRNISTIDF 57
 977 ELDDKDKKL--DEENELDDKK--KGLDENELDDKDKGLDENELDDKDKKL--DE 1030
 58 ENK--DLKLI-----KKKFRVDDFTSETGRMEBYDYKDDKKNIIAYDGT 104
 1031 ENELDDKDKKLDEENELDDKKGLDENELLEBRKKQKMDENILDEKKKEIYVNDHT 1090
 105 DLEYETGLDEIKSI--YGLSPS--KDGIF-----EILGKISNYSKNAK----- 146
 1091 FIQTEHNLNNEKTKIQDYNIIEALEIKEMSKKILAEETIKIKIENDENIDKRSQIEN 1150
 147 VYGNNGYSIEIKATKYDFHSGKTMTFDLYANINDIVDGLAFAGDMRLFVKNDQKKAIR 206
 1151 IYRRNSMMIDNINSYSKSNVKTFTFNDNIEE----- 1184
 207 IRPEKIKETSEYVYSSYGVNIELGSDLSKNKPDNLTQMSGKIYDSEKQVYLLKD 266
 1185 ---EDKIQOKNTKTYTDISNM-----LTKNNKSSIYINSSEKQIINEQEIYSKN 1234
 267 NI-----ILRKGYALKTTYNPKGTDMEGNYVSK-----EDIAKIQ-AAPN 309
 1235 NTFNNDIEKDLTNWKKNNNNNNISYKKQSNIVDINDSNKFLMDTSSIIKIQNDINEK 1294
 310 LRALS--ETTIYADSRNVEDGRSTQVLM--SALDGRNIIRYQVTFKNDGKGAIDKQGN 366
 1295 VKLLKQREIYEEKCNFEKKNELHLKENVLANKNIIRK--ENELNKEKEEELKKQKE 1352


```

QY 367 LVTDSKLVLF-----GKDKKEY-----TGEDKFNVAIKEDGSMFLPIDTKPVNL5W 413
DB 1353 ILLSKENIINYSLNSSKINDKLEYNKLEBVTKKGQIINSTM5KFSHL-SDTN--NIQ 1409
QY 414 DKRYFSPSSANKTYVNPPEFYLRGKISDKGCFNWEIARV-----NESVDNLYLY-G 463
DB 1410 FKQNDI---NDKLYENISSY--GNMLSHDNMKYDNNLSGDNM5HDNMNARDNICH5N 1465
QY 464 DLHIDTRDPIIKLVN-----KQDIDMDGMQDYKANGFPDVTMDGAVY----- 509
DB 1466 NISYDNPSPYNTMKLSLEBENNENKKNIDY---YDKDNFVDNNIINHNFNISKI 1521
QY 510 ---LQTSYDLNAKAVGVHYQFLYDVKPEVNIDPKGNTSIEYADGKSVFVINDKRNNG 566
DB 1522 SNDSISGNMDNR-----LHDNL-----NMSLKEIEMVKSILKSRDEIINT 1562
QY 567 FGEIOEQ-----HIYINGK-----BY-TSPNDIKQIIDXTLN 598
DB 1563 LKLIKIKQTEEEKKONNINQINDISSDMNIIYINGNISGVNYDTSMHVLPEPKKFLN 1622
QY 599 IKIVVDFAFNTVKEFIINKDQ---EVSLEKPHAVTYTIOGKEMSSITVSEBDFI-- 653
DB 1623 LS-EVED---NEKMKISLSEIWSLYKEIINNFK-----EENINIVLKKNEFIQ 1667
QY 654 -LPVYKGELEKGYQFGMEISGEGKKAGYVINTSKDPTFKPVFKIEKKEENKPTF 712
DB 1668 LLLNFINDLRNRYL---KENYKEKGVH5QIISERREDIKELQNTLNEKJKE----- 1719
QY 713 DVSKKQDNPOVNH5QLNESHK---KEDLQREH5QKSDSTKDTATVLDKNNISKST 767
DB 1720 -ISYKMKLLKMN--QINDTYKLNKRRSLSTVELLKQ--DIKFLNEDVLLKKEKAVTLST 1772

RESULT 14
QY 09YVT6 PRELIMINARY; PRT; 1127 AA.
AC 09YVT6;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein MSV156.
GN Name=MSV156;
OS Melanoplus sanguinipes entomopoxvirus (MSPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=83191;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=9102612; PubMed=9847359;
RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus.";
DR EMBL; AF063866; AAC97677.1; -.
DR J. VIROL. 73:533-552(1999).
DR PIR; T28317; T28317.
KW Hypothetical protein.
SQ SEQUENCE 1127 AA; 134265 MW; F185DA1D5A3FE7D1 CRC64;

Query Match 6.3%; Score 253.5; DB 2; Length 1127;
Best Local Similarity 21.5%; Pred. No. 0.0098;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;

```

```

QY 162 KYDFHSKWTPLFYANINDIVDLAFAGDMRLFVK--DNDOKKA-EIKIRMEKIKET-- 216
DB 368 LPDNDIQKLANDITBQNNKITD--FPNNSTRIFKEGLDIEYKKJDIIDKNNNQKLEESYK 425
QY 217 ---KSEY---PYVSSGVNVIIEGGLSKNKPDIITKSESKTYSDSKQOYLKXNTI 269
DB 426 KIDQETEVYKKNIKNEVYNDIIEKNNNIIQKLEENENKIKIDQETEVYKKNIKNEY--NDIIE 483
QY 270 LKGYALKTQTTNPG-----KTDMLEGNVYSKEDIKAKIQKANPILRALSETTYAD 321
DB 484 LKNNNLOKLEENKINDKLYLKNDIESNTELFNGLNLSIDPKDSRELAKLN--TEYSQ 541
QY 322 SRN--VEDRSTQSVL-----MSALDGNNIIRYQVFTFRMDKGAIPROGNLYTDSK 373
DB 542 LKQDLLENINKNTBELMKLSDNKLSLE-----QLYDSKNIL-DGIDIKIINSLAKEN- 592
QY 374 LVLFQKDKKEYTGEDKFNVAIKED--GSMLEFIDTKPVNL5MDKNFY--PSKSNKIT 427
DB 593 ---DKIDYFNNIEKFIDYVNIENKFNGLNLSIINKIINNDOFKYINSKIDSKE5LS 648
QY 428 VNPPEFY-----LRGKISDKGCFNWEIARVNESVDNLYLYGDLHIDNTD 472
DB 649 TWFDDIFNAKQIASITNNIENISNKKIKD-----INEF1ISNEDSSKEL-LDEIRK 698
QY 473 FNIKLVKQGDIDMDGMQDYKANGFPDVTMDGAVY-LQTSYDLNAKA-----V 522
DB 699 YKQQPD-KIKDAMNTEVSFE--NTLOKQIDSKSINELTNAVYDIINTQANDLDDGLANN 756
QY 523 GVHYQFLYDVKP-----EVNIDPKGNTSIEYAD--GKSVFVINDKRNNGFGEIOEQHI 576
DB 757 GSEFKLVYNNASDLDLDTIQKNNDEKQVNEYLEKKNQ5SEINDIVNN----- 805
QY 577 YINGKRYTSFNDIKQIIDITLNIKIYVQDPAARTYKEFIINKDGEVSE---LKHRYVT 633
DB 806 FL--KELIKFNTE--TNKSLN-ELLTND--DINDKFKLYELKIKSTNNILKIKYK-- 855
QY 634 VTIOGKEMSSITVSEBDFILPVYKGELEKGYQFGMEISGEGKKAGYVINTSKDPT 692
DB 856 NEIDVNEGLSIVIELOPINSFLETFQO-----STSH-----INFLANTLA 900
QY 693 -IKPVFKIEKKEENKPTFVSKKQDNPOVNH5QLNESHKEDLQREH5QKSDSTYD 751
DB 901 GINDVLYNKLNLIKIMADTTRRGDTNR-----DEIKQI5SENIK5QKFNKKE--XD 950
QY 752 VTATV-----LQKNNISKSTTN 770
DB 951 LKKLISFNDKLNKNYISAGYTEYN 974

RESULT 15
QY 08IIT21 PRELIMINARY; PRT; 2227 AA.
AC 08IIT21;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PF11_0354;
OS Plasmodium falciparum (Isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=2225705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Sub B., Peterson J., Angioli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaideya A.B.,
RA Martin D.M., Fairclark A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carnuci D.J., Hoffman S.L., Newbold C., Davis R.W.,

```

RA Fraser C.M., Bartell B.:
 RT "Genome sequence of the human malaria parasite Plasmodium
 falciparum."
 RL Nature 419:498-511(2002).
 DR EMBL: AB014841; AAN35938.1; -
 HSSP: 0931C2; 1MWT.
 KW Hypothetical protein.
 SQ SEQUENCE 2227 AA; 262841 MW; 6D5D4BEEF643339C CRC64;

Query Match 6.3%; Score 252.5; DB 2; Length 2227;
 Best Local Similarity 19.1%; Pred. No. 0.024;
 Matches 190; Conservative 155; Mismatches 319; Indels 329; Gaps 49;

QY 10 FKVLNGKE-----GSLKDTTGVH-----HQENESIKYK 42
 DB 630 YNNNNNNKCTCTINIKSENKYPFKSHVNSYMKVTNHLPHRNAITSNNRNNEEYKEX 749
 QY 43 S6FTIDRNIST-----IDFENK---DKKLI-----KKKREVDFT-----SE 79
 DB 750 EK--DRNITNGNNYLVAYNNSCIPPLKRMIPIDGVNKSINKLNVNTTQRTSSVSY 806
 QY 80 TGRMEEDY-----KYDDKNIAYDGTLEETEKLDEIKSIYGLSPSK 128
 DB 807 TKNIDENSFDMPIINGIRESKYISNNNI---NGNSIGFNSKLDNYHQSMNV----- 858
 QY 129 DGHFILLKISVNSKAKVYGNNTYSIEIKATKYDFHSKTMTP-----DIYANIN-- 179
 DB 859 -----NESYPLKMMKNY-----IEHYDDKNNIPLVKNYEDTYSNIHNGI 900
 QY 180 -----DIVDGLAFAGDMR-----LFFKDN---DOKK----- 202
 DB 901 HENSMIKYNTLKKACTFHGYSRHHQKMTSENLINQKQNSHYHNGTGLKPLVNTNN 960
 QY 203 -----ABIKIMPEKIKETKS---EYVSSSYGNVIELEGDLKKNKPDNLTKMESGKI 253
 DB 961 VAVNEFADINLSAQKRHLHLSKMGYEDKEMENYRNKITYNNINNNNNNDN-----NI 1013
 QY 254 YSDSEKQY---LLKDNITLRKGYAL---KVTYNNPEKTMLENGYYSKEDIK--- 302
 DB 1014 YNDNEYCQYNNNSYCFDHSDLKMMFPLNHQNSKLTHTSHNNKNSFFNGINVESKHLANPEI 1073
 QY 303 -----IQKANP-----N 309
 DB 1074 KTFANNSYPIINQGLINCPLOCLGYDSNQKKNHNVYIKKNEYLANKIGIINVLKREG 1133
 QY 310 LRAALS-----ETTYADSRNVEDGRSTOSVLSALDGFNIIRYOVFTFKMD 356
 DB 1134 LRKISTHNGKPEPSFNMGMKNVYMGGLNTQDNVNNNNNNKESC-DNIKHMRTKSLNFVSR 1192
 QY 357 K-GE--AIDKGNLYTDSKLVLPGR-DKKEYTGEDKFNVEAIKED-GSMLFID---TKP 408
 DB 1193 SYGEHKSILVYOECYVKNKLI--NKVNDKCY--EDNNNSYLNEDDNASMQFYEBTNSNP 1248
 QY 409 VNLAMD---KYNFNSKSNKIYVRNPEFLRGKISDKGCFNME-----LRVNESV----- 455
 DB 1249 YIVDQENNMKNYV---NVLNNNSNTYVDSKNTYDKSKENAKENKSDTLNNEINTHTLKD 1304
 QY 456 -----VDNYLLYGDHLIDNTRDFNIKLVNVDGDIWDV--GMKDYKANGFPDKV--TDMG 506
 DB 1305 QKKKIQNNNEFISEQADIENINSQEEVEYKEHEPL-WVINASNEKKSYEELIYSDMS 1363
 QY 507 NVYLTGTGSDLAQAVGVAHQFLYDNVKEVNIIDPKGNTSIEYADGKSVFENINDKRNNG 566
 DB 1364 NRVTKRKTSDMNNVEVLNEDMLLTTEKYKQLEKENKIMDYETVEENINTIKENTVD 1423
 QY 567 PGGEIOEQHIYINGKEYSFNDIKQIDITLNIKIYVVDPAKNT--TVK--EPIINKDNG 622
 DB 1424 INEEVR-----NEQGRRESININD-----TNINHTIDEPDYTNFIIDIECVANNEN 1472
 QY 623 EYSELKPHRVYTLIONGKEMSSSTIVSEEDFILPVYKGELEKGYQFPDWEISGFEGKQDAG 682
 DB 1473 MWNSTEQYTFYHDTNRNHLVDK--NNQNF-----FEEBGLNEALNFEKK--V 1516

QY 683 YVINISKD-----TFIKPVFKIIEKKEBENKPTF-----DYSKKRON 720
 DB 1517 YIENNTKDXDHKGDSKTSNLTSLRNTICSENDHNEKNENTYVVRKGEKGIKRVSMKKRN 1576
 QY 721 PQVN-HSQLN-----ESHKKEDLQREBHSOKS 746
 DB 1577 EKLNEENIYNNIYDKMDVHRQNDITKENDEEN 1609

Search completed: February 16, 2005, 19:19:52
 Job time : 188 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:11:30 ; Search time 44 Seconds
(without alignments)
1690.354 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGEIASEKFKNLGNGKEGS.....ATVLDKNNISKSTNNPK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	2140	2 F95074	serine proteinase,
2	3709.5	92.1	2144	2 A97942	metalloproteinase
3	253.5	6.3	1127	2 T28317	ORF MSV156 hypote
4	250	6.2	1650	2 T18444	hypothetical prote
5	235.5	5.8	3724	2 T28427	hypothetical prote
6	231.5	5.8	2269	2 T28677	rhodopsin protein -
7	230	5.7	2401	2 T28676	rhodopsin protein -
8	228	5.7	4550	2 T18440	hypothetical prote
9	221.5	5.5	1711	2 T18429	hypothetical prote
10	219	5.4	1640	2 A24594	probable major sur
11	218	5.4	1631	1 SAZOK1	major merozoite su
12	218	5.4	1639	1 S05603	major merozoite su
13	217.5	5.4	1546	2 G90603	reticulocyte-bindi
14	215.5	5.4	1252	2 B42771	DNA-directed RNA p
15	214	5.3	2339	2 A45597	membrane nucleas,
16	212.5	5.3	1125	2 E90598	hypothetical prote
17	212.5	5.3	2166	2 G70163	serine/threonine-s
18	212.5	5.3	2485	1 H71621	protein with 5'-3'
19	212	5.3	1188	2 A71621	microbial collagen
20	211.5	5.3	1104	1 A36866	REBA-H3 antigen pr
21	209	5.2	1558	2 B71603	hypothetical prote
22	207.5	5.2	4688	2 B82885	chromosome segrega
23	207	5.1	1141	2 B97116	hypothetical prote
24	207	5.1	1191	2 B97116	hypothetical prote
25	207	5.1	1191	2 C71622	probable cell surf
26	206.5	5.1	1622	2 A81717	mature-parasite-in
27	206	5.1	1526	2 A45605	RAD2 endonuclease
28	204.5	5.1	1516	2 E71619	hypothetical prote
29	204.5	5.1	2500	2 G71609	

30	204	5.1	1245	2 D71613	GAF domain protein
31	203.5	5.1	3394	2 T18501	hypothetical prote
32	203	5.0	1166	2 T28680	fibrogen-binding
33	203	5.0	4981	2 T18489	hypothetical prote
34	202.5	5.0	763	2 G97026	superfamily 1 DNA
35	202.5	5.0	1939	2 T18372	repeat organellar
36	201	5.0	1712	2 C71618	hypothetical prote
37	199	4.9	888	2 E71608	ATP-dept. acyl-CoA
38	199	4.9	980	2 E71606	hypothetical prote
39	197	4.9	1386	2 AC1533	surface protein (L
40	196.5	4.9	2523	2 T18477	hypothetical prote
41	196	4.9	1714	2 E71609	Ser/Thr protein Ki
42	195.5	4.9	2380	2 E71604	hypothetical prote
43	195	4.8	1008	2 T18508	hypothetical prote
44	195	4.8	1302	1 JG5009	surface-located me
45	195	4.8	1365	2 T30822	Imp1 protein - Myc

ALIGNMENTS

RESULT 1	
F95074	serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae	
C:Date: 03-Aug-2001	#sequence_revision 03-Aug-2001 #ext_change 09-Jul-2004
C:Accession: F95074	
R:Retcelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hel-	
son, J.D.; Umeyam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Kadane, D.; Holtzapple,	
nson, T.; Hickey, E.K.; Holt, I.E.	
Science 293, 498-506, 2001	
A:Authors: Loftus, B.D.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,	
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.	
A:Reference number: A95000; MUID:21357209; PMID:11463916	
A:Accession: F95074	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-2140 <KUN>	
A:Cross-references: UNIPROT:Q97RV6; GB:AE005672; PIDN:AAK74791.1; PID:g14972117; GSPDB:	
A:Experimental source: strain TIGR4	
C:Genetics:	
A:Gene: SP0641	
Query Match	100.0%; Score 4026; DB 2; Length 2140;
Best Local Similarity	100.0%; Pred. No. 1.9e-164;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 KLGEIASEKFKNLGNGKEGSLKDDTGVVHHQNEBESIKKSSFTIDRNIISTIRDFENK 60
DB	1334 KLGEIASEKFKNLGNGKEGSLKDDTGVVHHQNEBESIKKSSFTIDRNIISTIRDFENK 1393
QY	61 DLKULIKKKFRVDDFTSTGKRMEEYKYDDKGNIIAYDDGTLEYTEKLDLDEIKSI 120
DB	1394 DLKULIKKKFRVDDFTSTGKRMEEYKYDDKGNIIAYDDGTLEYTEKLDLDEIKSI 1453
QY	121 YGVLSPSKDGHEIIGKISNVSKNAKVYGNKYKSEIKATKYDFSKMTDFLYANIND 180
DB	1454 YGVLSPSKDGHEIIGKISNVSKNAKVYGNKYKSEIKATKYDFSKMTDFLYANIND 1513
QY	181 IYDGLAFAGDMELFYKNDQKKAIEIKIRMEKIKETKSEYPVSSYGVNIELEGEDLSKN 240
DB	1514 IYDGLAFAGDMELFYKNDQKKAIEIKIRMEKIKETKSEYPVSSYGVNIELEGEDLSKN 1573
QY	241 KPDNLTAKMSKTIYDSSEKQVLLDNITLRGVALKVTYTPGKTDMEGNGVSKKEI 300
DB	1574 KPDNLTAKMSKTIYDSSEKQVLLDNITLRGVALKVTYTPGKTDMEGNGVSKKEI 1633
QY	301 AKIQKAPVLRALSETTITVADSRNVEDGRTQSVLMSALDGNIIIRYQVFTFKMDKGBA 360
DB	1634 AKIQKAPVLRALSETTITVADSRNVEDGRTQSVLMSALDGNIIIRYQVFTFKMDKGBA 1693
QY	361 IDKDGMLVTDSSKLVLPKDDKEYTGEDKFNVKAIKEDGSMFLPDTKPVNLSMDKXVFP 420

Db 1694 IDKGNLVTDSKVLFGKDDKEVTGDEKFNVEAIKEDGSMFLITKPVNLSMDKNYENP 1753
Qy 421 SKSNKIYVRNBEFYLRGKISDKGFWNMLRVNESVVDNYLIYGDHLIDNTRDFNKLNVK 480
Db 1754 SKSNKIYVRNBEFYLRGKISDKGFWNMLRVNESVVDNYLIYGDHLIDNTRDFNKLNVK 1813
Qy 481 DGDIMDMCKQYKANGFPDKXTDMDGNVYLQYGSIDLAKAVGVHYQFLYNNVKEVNI 540
Db 1814 DGDIMDMCKQYKANGFPDKXTDMDGNVYLQYGSIDLAKAVGVHYQFLYNNVKEVNI 1873
Qy 541 PKGNISIEYADGKSVFVFNINDKRNNGFDEIOEHIYNGKEYTSFNDIKQIIDTLNLIK 600
Db 1874 PKGNISIEYADGKSVFVFNINDKRNNGFDEIOEHIYNGKEYTSFNDIKQIIDTLNLIK 1933
Qy 601 IVVKDFANNTVKEFLINKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 660
Db 1934 IVVKDFANNTVKEFLINKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 1993
Qy 661 LEKGYPFGWEISGEGKDDAGVYNLSKDTPIKVPFKIEEKEENKPTFDVSKKDN 720
Db 1994 LEKGYPFGWEISGEGKDDAGVYNLSKDTPIKVPFKIEEKEENKPTFDVSKKDN 2053
Qy 721 POWNHSQLNESHKEDLQREHSHQSDSTKDVATVLDKNNISSTKSTNNPNK 773
Db 2054 POWNHSQLNESHKEDLQREHSHQSDSTKDVATVLDKNNISSTKSTNNPNK 2106

RESULT 2

A97942 metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
R:Accession: A97942
R:Host: R.Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; Dehoff, B.S.; H
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
Y: Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaakunas, S.R.;
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97942; MUID:21429245; PMID:11544234
A:Accession: A97942
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2144 <KUR>
A:Cross-references: UNIPROT:Q8DOP7; GB:AE007317; PIDN:AAK9365.1; PID:g15458138; GSPDB:C
A:Genetics:
C:Keywords: hydrolase; serine proteinase

Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
Best local Similarity 92.4%; Pred. No. 6,4e-151;
Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;
Qy 1 KLGEIAESKFNKLGEGSLKQDTTGVHHQENESIKESPTIDRNISTIRDFENK 60
Db 1339 KLGEISBKLKAKSPRENT-NNNAQKEDSKPDKSVESGASLEIKTISTIRDFENK 1397
Qy 61 DLKLIKKKPREVDDFTSETGKMEYKYDDKNNIAYDGTDLBXYETEKDLDEIKSI 120
Db 1398 DLKLIKKKPREVDDFTSETGKRIREYDYKYDDKNNIAYDGSALQYETEFDEIKSKI 1457
Qy 121 YGVUSPSDGHFEILGKISNYSKNAKYVGNNTKSIETKATKYDHSKTMFTDLYANIND 180
Db 1458 YGVUSPSDGHFEILGKISNYSKNAKYVGNNTKSIETKATKYDHSKTMFTDLYANIND 1517
Qy 181 IVDGLAFAGDMRLFYVDNDOKKALIKIMPEKIKETSEYVSVSYGVNIEIGEGDLSEN 240
Db 1518 IVDGLAFAGDMRLFYVDNDOKKALIKIMPEKIKETSEYVSVSYGVNIEIGEGDLSEN 1577
Qy 241 KPDLNLTAKESGKIYSDSKQOYLKDNILIRKGYALKVTTYNPGKIDMLENGNGVSKEDI 300
Db 1578 KPDLNLTAKESGKIYSDSKQOYLKDNILIRKGYALKVTTYNPGKIDMLENGNGVSKEDI 1637

Qy 301 AKIQANPNLALSTTTIYADSRNVEDGRSTQSVLMSLDGDFNITRYVFFPKANDKGEA 360
Db 1638 AKIQANPNLALSTTTIYADSRNVEDGRSTQSVLMSLDGDFNITRYVFFPKANDKGEA 1697
Qy 361 IDKGNLVTDSKVLFGKDDKEVTGDEKFNVEAIKEDGSMFLITKPVNLSMDKNYENP 420
Db 1698 IDKGNLVTDSKVLFGKDDKEVTGDEKFNVEAIKEDGSMFLITKPVNLSMDKNYENP 1757
Qy 421 SKSNKIYVRNBEFYLRGKISDKGFWNMLRVNESVVDNYLIYGDHLIDNTRDFNKLNVK 480
Db 1758 SKSNKIYVRNBEFYLRGKISDKGFWNMLRVNESVVDNYLIYGDHLIDNTRDFNKLNVK 1817
Qy 481 DGDIMDMCKQYKANGFPDKXTDMDGNVYLQYGSIDLAKAVGVHYQFLYNNVKEVNI 540
Db 1818 DGDIMDMCKQYKANGFPDKXTDMDGNVYLQYGSIDLAKAVGVHYQFLYNNVKEVNI 1877
Qy 541 PKGNISIEYADGKSVFVFNINDKRNNGFDEIOEHIYNGKEYTSFNDIKQIIDTLNLIK 600
Db 1878 PKGNISIEYADGKSVFVFNINDKRNNGFDEIOEHIYNGKEYTSFNDIKQIIDTLNLIK 1937
Qy 601 IVVKDFANNTVKEFLINKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 660
Db 1938 IVVKDFANNTVKEFLINKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 1997
Qy 661 LEKGYPFGWEISGEGKDDAGVYNLSKDTPIKVPFKIEEKEENKPTFDVSKKDN 720
Db 1998 LEKGYPFGWEISGEGKDDAGVYNLSKDTPIKVPFKIEEKEENKPTFDVSKKDN 2057
Qy 721 POWNHSQLNESHKEDLQREHSHQSDSTKDVATVLDKNNISSTKSTNNPNK 773
Db 2058 POWNHSQLNESHKEDLQREHSHQSDSTKDVATVLDKNNISSTKSTNNPNK 2110

RESULT 3

T28317 ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
R:Accession: T28317
R:Host: R.Alonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A>Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:99102612; PMID:9847359
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: UNIPROT:Q9YVT6; EMBL:AF063866; NID:g4049647; PIDN:AAK97677.1; PID:g4
C:Genetics:
A:Note: MSV156

Query Match 6.3%; Score 253.5; DB 2; Length 1127;
Best local Similarity 21.5%; Pred. No. 0.0015;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;
Qy 7 ESKRNLNANGEGSLKQDTTGVHHQENESIK-ESKSTTID-RNISTIRDFENKOLK 63
Db 196 EIEFNDIDNVKSEINKQDELKILDESKKEFIKQEBELINTDIPKQBELKKNDEKIN 255
Qy 64 KLIKKKPREVDDFTSETGKME-----EYDYKYDDKNNIAYDGT--DLEYETE 111
Db 256 FNIDBKQKLDQINSKINTLNTENIKGVNLTETKKNKSNLQNELNKDSTIKSLDEKQK 315
Qy 112 KLDEIKSKIVGLSPSDGHFEILGKISNYSK-----NAKVYGNNTKSIETKAT 161
Db 316 LLDDELKNNINNTISLYNKS-----TKITNIOQLLESSTLDENNNI-----NINELKSKIX 367
Qy 162 KYDHSKTMFTDLYANINDIYDGLAFAGDMFLFYK--DNQKKA-EIKIMPEKIKET-- 216
Db 368 LFDNDIOKLDNDITEQNNKIDT--FFNNSTRIIFPEKIDTEYKCKIIDIDIGNNMLQKLEESYK 425
Qy 217 -----KSEY---PYVSYGVNIEIGEGDLSEKPKPNLTAKESGKIYSDSKQOYLKDNII 269

426 KIDQTEYKKNKINKEVNDIIEIKNNNLOKLEENKIDQTEYKKNKINKEV--NDIIE 483
 QY 270 LKRGVALKVTYTPG-----KIDMLENGVSKEDIKQKPNRALSFTIYAD 321
 Db 484 LKNNLOKLEENKINNDKLTLYKNDIESNTLFTKLNISDFDKSREIAKLN--TEYEQ 541
 QY 322 SRN--VEDGRSTQSVL-----MSALDGFNIIRYQVFTFRMDKGEAIDKGNLYTDSK 373
 Db 542 LKRDLEINIKNTKMLKSDNKLSSLE-----QLYDSKKNIL--DGIDKIVNSLEKKN- 592
 QY 374 LVLFGKDDKYEYGEDKFNVAEKED--GSMFLIDTKPVNLSMDKNYFN--PSKSNKTY 427
 Db 593 ---DKIDYFNSNIEKFDIYNVLENKFINLDSIINKIINNDFKEXYINSKIDSKSNES 648
 QY 428 VSNPEFY-----LRGKISDKGFMWELRVNESVVDVYLLYGDHIDNTED 472
 Db 649 TWFDDTFNAKQOASTNNIENISNKIKD-----LNEFTISNEDSSKEL--LDEIRK 698
 QY 473 FNIKLVKXGDIDMDGKDYKANGFPDKYTDMDGNY--YLQGYSDLNAKA-----V 522
 Db 699 YKQGPD-KIKDAMNTEVKSFE--NTLOKDIDISKSNINELNADINTKANDLDDKLNKY 756
 QY 523 GYHYQFLYNNKPE-----EVNIDPKGNTSIEYAD--GKSVYFNINDKRNNGFDEIOEQHI 576
 Db 757 GSEFKNLYNNASDLDTTIQKNDKVKOLNEYLEKKNKQSIIEINDIVN----- 805
 QY 577 YINGKRYTSFNDIKOIIDTKINIKIYVCDPARVTVKEPILNKDQGEVSE--LKPHRYT 633
 Db 806 FY--KELIKFNNTTE--TNNSLN--ELLTND--DINDIKFLYKELNKISTNNILKTYK-- 855
 QY 634 VTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGMEISGPEKKDAGVINTKOTF- 692
 Db 856 NEIDNVNEKLSIYENLOFINSLFSLIEFNQ-----SITSH-----INFLANTLA 900
 QY 693 -IKPVFKLIEKKEENKFTFDVSKKKNPQVNSQLNESHKREDLQREHSGKSTGSD 751
 Db 901 GINDVNLKNTLKIIMDPTTRGDTNIR-----DEIKNOISSENIKSKOFNEKKN- 950
 QY 752 VTRATV-----LDKNISKSSTNN 770
 Db 951 LKKLISFNDKLNKNISAGYTEYN 974

RESULT 4

T18444
 Hypothetical protein C0385c - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T18444
 R:Lawson, D.; Bowman, S.; Barrell, B.
 submitted to the EMBL Data Library, August 1997
 A:Reference number: Z18935
 A:Accession: T18444
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1650 <LAW>
 A:Cross-references: UNIPROT:O77328; EMBL:Z28547; NID:e1325376; PID:e1427940; PIDN:CAB111
 A:Genetic: 3
 A:Map position: 3
 A:Introns: 1597/3; 1625/3
 A>Note: C0385c

Query Match 6.2%; Score 250; DB 2; Length 1650;

Best Local Similarity 20.8%; Pred. No. 0.0034; Indels 328; Gaps 48;

Matches 205; Conservative 134; Mismatches 320; Indels 328; Gaps 48;

QY 20 SLKQDTGVYHHQNEESIKESSTIDRINISTIRDFENKDKLKKKKFRE---VDPF 76
 Db 394 SLWENNEKRIYKKEEYNSKNEV-----HFKND--DSGKKKNSSECLDQ 443
 QY 77 TSETGR--MEYDYKDKGNITAY--DDGDL--YETKLDKSKIKYIVLSPSKD 129
 Db 444 KKTYKTYTIEQRYVFNDRNNNAVYKIDTHKKEGYLNMIVQSEYKYG--SNMK 501

QY 130 GHFEILGKIS-----NVEKNAKVVYGNVYKSI--EIKATKYDFHSTWTFDLYANINDV 182
 Db 502 DEMETIYQHTNDFNINENLNKKIYF--DYEGDPEKKKKKLDHITQOKEKKNINDL 560
 QY 183 DGLAPAGWRLFVKNDQKAEIKIRMPKIK--ETKSEYPVYSSYGNVIELEG- 235
 Db 561 K-----DHLNDKETEKNKEIEEEKKNKIEIEEEKKNKIEIEEEKKNKIEIEEEKK 614
 QY 236 ---DLSKKPDLUTMGESGKIYSDEKQOYLLKNDI-----LR 271
 Db 615 KKEIEEEKKKKI--EMEEKKNKIDKKNYANDKIISHIDVNCNKKIDALLHIEEK 673
 QY 272 KGVALKVTYTPGKT-----MLENGVSKEDIK--IQKPNL-- 310
 Db 674 KTGHEINLYKEIKRYQKMLNDENSIMLEHKKNTHTQVNNNLDDTKMLQKELILN 733
 QY 311 RALSETTYADSRNVEDGRSTQSVLSALDGFNIIRYQVFTFRMDKGEAIDKGNLYT 370
 Db 734 NDKKTFILSKSKNI-----TSNVLSSKIPG-----TLSTKGLNATIKTIKKD-- 779
 QY 371 SKGLVLFQ-----KODKE--YTGEDK--NYEA-----IKEDSMLFIDTK 407
 Db 780 NEKKKYVHDHRRKONITKKNKEFINIYKGRNANVEIGSEVCNKNVYAGDNNKAVENK 839
 QY 408 P-----VNLSDKNYF-----NPSKKNKIYYRANPEF 433
 Db 840 QGDDNNMAYENKQGBENNVIVANIERSSSFPFTHRRKNITSTSDTKAKNEQIKYPHL 899
 QY 434 YLRGKISDKGFMWELRVNESVDNY-----LIYGLDLDN----- 469
 Db 900 Y--SNKNEDEKDKSIFLKININENIKKQYKQEKISTLEKKVFVKQDNVITNNDKHTSSK 958
 QY 470 -TRDENI-----KLN----- 478
 Db 959 INDDFNITIDKQKGLNNPVDLNRIKNEYKILEKDSHMSKIONNLMKKKTNPPTNK 1018
 QY 479 -VKDGDIDMGKMDYKANGFPDKYTDMDGNVYL-----QTG 513
 Db 1019 GISSSTISSSTKRNKDCGIEKKNLSLKYTCIRKKNVDSIKLNDKADLYKDKKTS 1078
 QY 514 YSDLNKAVGAYHYQFLYDNVKEVINIDPKGNTSIEYADGKSVFN--INDKRNNGFDG 569
 Db 1079 FNDINRAKGNFK-----KSDY--PKKNMVDVTKGKR--VFNVPVTLNNYRNN-- 1124
 QY 570 EIQEGHIYINGKRYTSFNDIKOIIDTKINIKIYVCDPARVTVKEPILNKDQGEVSELP 629
 Db 1125 -----YTRSNK-----NNVK-----NGKRV-----GIKKIVLKE--KOKSLHP 1156
 QY 630 HRVTVTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGMEISGPEKKDAGVINTLSK 689
 Db 1157 EGVEADKGLNSYNDKYLEKGFQDINEBEK-----YKNNK--KKYKI--K 1201
 QY 690 DFFIKPVFKLIEKKEENKFTFDVSKKKNPQVNSQLNESHKREDLQRE--HSQKS 746
 Db 1202 SNSIPIIKIERKSDNDN--DNKKNNDINNSNNINSNDKCLFLSKEDRVHLKNN 1258
 QY 747 DSTKQVTAIVLDKNNISKSSTNNPNK 773
 Db 1259 NIIVNTMMFRKQNSCDNNTTSLNKK 1285

RESULT 5

T18427
 Hypothetical protein C033c - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T18427
 R:Lawson, D.; Bowman, S.; Barrell, B.
 submitted to the EMBL Data Library, August 1997
 A:Reference number: Z18935
 A:Accession: T18427
 A:Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA
A.Residues: 1-3124 <LAN>
A.Cross-references: UNIPROT:Q07320; EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB111
C.Genes:
A.Introns: 307/1; 1545/2
A.Note: C0335C

Query Match 5.8%; Score 235.5; DB 2; Length 3724;
Best Local Similarity 18.7%; Pred. No. 0.041;
Matches 192; Conservative 150; Mismatches 312; Indels 371; Gaps 45;

5 IAESKPKLNGKES--LKK-----DTGCVHHQNEESTK-----EKSSPT 46
249 IPKNLEPNLNKQNDYQNRNIIIMDVNDINPLHPBQONSLNKNKCLTGTNKEKYM 308
47 I-----DRNISTIRFENDKLKLKKKREYVDFSETGKMEEDYKXDKG 95
309 IPKGYEODKEKESLITINQNDKKYKKK--KKKXSELQDSNINNTLTSKRYT--- 363
96 NIAYDGTDEYETFEKLEDEKSKYGVLSPEKDGHEILGKISVSKAKYVYGNNTYS 155
364 -----CGMDKETETDEQNR-----RNTQVASICVAVETEKQKKTNKK 407
156 IEIKATKVDHFKMTFDLYANINDIYGLAFAGDMR--LFVQNDQKKAELIKIMP--EK 212
408 ---SVQKODLVEENALD---NL-IITDINFDVYTKCKITIDNNENNIDNNIYENNK 460
213 IKETSEYVYVSSYG-NVI-----ELGEGDLSKNKPNMLTQMESGK 252
461 LKQDQS-YDLFSEBEGKNLILGVNBESEFEYFENIEKLGQRKQKQKNDK-----T 512
253 IYSDSEKQOYLKKNIIILKGYALKVTTYNPGKTMLEBNGVYSKEDIKQKAN-PNLR 311
513 IYNNNEEQDDLNRI-----NKIESINNNDN---NNNNNNKKEFKIKTEHILNKE 562
312 ALSETTIYADSRNVDSRSTQSVLMSALDGNIIYRYQVTFPMQNKGAIDGMVLYDS 371
563 SISKIKSSPKRKE-----IKKLTYTNQEDSTELKKELEITNN 604
372 SKLVFGKDKETGEDKENVEAIKEDGMLFIDTKPVNLSDKNYFNPSSKN----- 424
605 KNNVY--BEDIGSNEDDEYIHVKE-----NLKEDABENNDKKNKNTKE 650
425 ---KIYRN-----PEFLRGKISDKGSENNELRVN--ESVDNLYLIGDLHIDNTRD 472
651 ILKSKVLENEKRTLEELRGK--NNIFKDEKNSLGEVYINIEIQINEBNKINDIOD 707
473 FNIK-----LVNVD---GOIMDKMDYANGPDPKVTDM----- 505
708 GNISKQKIIQSSSRNTDFNIDISLNDLLEKKEKKGKQHFIDMLVKADKNEISENINK 767
506 -----GNVYLQT-----GYSDLNKAAGVHY----- 526
766 ICDNNINNIYDESINNIYDESINNIYDESINNIYDENINNIYDENINNIYDE 827
527 -----QFLDYKPEVNID-----PKGNTSIF-----Y 549
828 NINNIYDEGINIKCDNILENNIKITNDIYQVEENNESIEKNEMLISLNDKINNNTNF 887
550 ADGKSVFN-----INDKKNNGPDEIOE----- 573
888 KENVDFINKIRGSELKIDKNIKQNNDDIYIMNDFIYENDTIINHKKEITNKEIDPLEI 947
574 -----OHIIYNGKETS---FNDIKQIDKTLNIIKIVVDPAFNTVKEFLNKDGE 623
948 NTQNEFIEHLIDIKKKKYTNDHFNPADKKFYE-NM-KILNKMCKKNKQDEPFTQBTGS 1005
624 VSELK-----PHRYVTYIIONGKMSSTIYSE 649
1006 LQSHKIKKKYKKEKHKDKNNBEKNILYDENQVSVLYSDHKIEDIODIHSIQNTICDE 1065
650 EDFILPVYKGELEKGYOPDGEWISGPEGKQAGYINLSKDFIIPVFKKIEKKEEENK 709

DB 1066 NN-IEQINESSKKVIRISGDM---ENKND-----HENKNDMEKK 1102
QY 710 PTFDVSKKQNDPOVNH-SQINESHRKEDLOREHSQKSDSTKDYATVLDKNNISKSST 768
DB 1103 N--DWEKNDNIRKQNDMEKKQNDMEKKQNDMEKKQNDME-----NENNNENKSDI 1153
QY 769 NNPNK 773
DB 1154 ENENK 1158

RESULT 6
728677
rhoptry protein - Plasmodium yoelii
C.Species: Plasmodium yoelii
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C.Accession: T28677; C45521
R.Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A.Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A.Reference number: Z20508; MUID:95021522; PMID:7935623
A.Accession: T28677
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-2269 <KE2>
A.Cross-references: UNIPROT:Q26223; EMBL:L27838; NID:g457145; PID:g457146; PIDN:AAA21304
R.Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A>Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple cc
A.Reference number: A45521; MUID:91101660; PMID:2270106
A.Accession: C45521
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 2131-2269 <KE2>
A.Cross-references: GB:M34283

Query Match 5.8%; Score 231.5; DB 2; Length 2269;
Best Local Similarity 19.7%; Pred. No. 0.032;
Matches 181; Conservative 147; Mismatches 330; Indels 259; Gaps 37;

37 ESISEKSFIDRNSTRDPENK-DLKKLKKKPRVVDPTSETGKMEY----- 87
109 EQLKELNKKIKDIKAIKRYKAYVLEKKELEKNNAYIDELANGSPRYVTGIENKNTIY 168
QY 88 -----DYRYDDKGNIIAYDGTDELEYETKLEDEIKSKIYGVLSPEKDGHF 132
169 NTIKSPQIYEGDIDTFYNELSSIVKEDPIDIEDKT-KLENLRSKIDYVDKIQKMEI 227
DB 133 E-ILKISNVSKNAKVYGNNTYSKIEIKATKYDFHSKMTFDLYANINDIYDGLAFAGDM 191
228 ETVKSHLNINISTNNKL-----PNTILEIKRYIYDEISK---ELNNMLDFKRNK--EKEL 276
QY 192 RLFFVQNDQKK--AEIKIRPME-----KIKETSEYVYVSSGVNIEIGE 234
277 SNKISDYKKGQQLSEYKSLERHNSQTNVDNTEBEAKQNVYDKSNEHMTIIPNE 336
DB 235 GDLK-----NKPNDLTQMS-----GKIYSD--SEKQYL-LKQNI---ILKGYAL 276
337 DEISKIIEVTKMDIELSKVNTYIDFNKRYKEFVNSHSQFTLETDKIKAEVSDKELK 396
QY 277 KATTYNPGKTMLEBNGVYSKE--DIKIQKANPRLALSET-----TIYAD-- 321
397 CEQSFNDKRSILNETKNSIEKEYOINTLKQVDEYIKVCKSYKSTIRFSSKQITLKML 456
DB 322 SRNVDSRSTQSVLMSALDGFNIRYQVFT-----FKNN-----DKGEAI----- 361
457 NQNTKVTETNSIDSKYIEKE-----QILTKQKTELENKFTFSLNHEANNELIKXFS 512
QY 362 DKDGNLVTSKVLFGDKDEYTGEDKFNVEAIKEDGMLFIDTKPVNLSDKNYFNPSS 421
513 DLKANLGINENMLYNGTEKEKTFND-----IKEKNIHINEISKIEIKHAIYVNS 566
QY 422 KSNKIYVNPFIYLRGKISDKGSENNELRVNESVVDNLYLIGDLHIDNTRDPNITLANK- 480

```

Db      567 BETEREIGINISLTKVFEK-----VKENVTLANKIKKYLKHYDFSGKEGNIKY 618
      481 -----DGDINDMGK--DYKANGFPD-----KTDNDGNN-----YLQGYSDLANK 520
      619 TDKIKKINDIDNAVSQDIOHINGLDIOKSSSESYSEMEQONKLEKVSNTFISDNVE 678
      521 AVGVHQLPYDWNKPEVNI-----DPKNTSIEYADGKSVFNIN-----560
      679 GKXKQOIIIVTKIDKKKNLYEELINKLSKIEKONTSLKVKDINLSYGNLGLFLE 738
      561 --DKRNGDFGEIOEHYIN-----GKEYTSFNDIKQIIDKTLN- IKI-----601
      739 QIDEEKKAKENTIKSMEAYIDLDNKKKSQEIETEMDIKMDINKEMELKISHDDKCC 798
      602 -----VKDFPANTVKEFLINDTGBVSELKPRVYVITONGKE 641
      799 HDKSNKHENISIDYDKSKKIIQDFRESKDIND-INKLQKNVSEQNNHNSDIN-----851
      642 MSSTVSEDFILPVYKGELEKGYQFDGWEISGFEKGKADAGVYINLSKDTFIKPVFKIE 701
      852 -----QCLNVAANIYILKANKIKKIIDKVK 877
      702 EKKEENKPTFVSKKKONPVNHSQNLNESHKEDLQREHSQKSDST---KDV-----T 753
      878 EYTSERIEK-----NKKINDLNNSEKVIKIEGDLSEKCRSKINSTLDDKOIDECIKN 932
      754 ATVLDKNNISKSTTN 770
      933 INVLLKNILNETNTN 949

```

RESULT 7

```

rhodtry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28676, A45521
R:Simla, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A>Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: Z20507, WUID:97077455, PMID:8920022
A:Accession: T28676
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: UNIPROT:Q26216, EMBL:U36927, NID:91041784, PID:91041785, PIDN:AA8412
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A>Title: Identification of the gene for a Plasmodium yoelii rhodtry protein. Multiple c
A:Reference number: A45521, WUID:91101660, PMID:2270106
A:Accession: A45521
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: GB:M34281

```

```

Query Match      5.7%; Score 230; DB 2; Length 2401;
Best Local Similarity 21.0%; Pred. No. 0.04;
Matches 187; Conservative 138; Mismatches 292; Indels 274; Gaps 46;

      2 LGBIASEKPNLNGKSGSLKKDTTGV-HHQENESIKKKSFTIDNRNISTIRDFENK 60
      589 IDKILYDKENITTDKTELETFGTSLNNHESNN---KELTYTVDLKANKGNKEN- 644
      61 DLKLLIKKKFREYVDFTSETGKMEBYKYDDKGNIIA-----YDGTDLLEYET 110
      645 ---MLYKQFNKSG-----KAVEDIKKKNVDINKIVSNIBITTYTSIYNEDTENEL 693
      111 EK-----LDEIKSKTYGVLSPEKDGHEFLGKISNVSNAKATYYGN 151
      694 GKSIELLNTKVLKVKANVTNINLEIKEL-----KDYDFDQFGK---EKNIKYPDEN 742

```

```

      152 NYKS-IEIKATYDPSHKTMFPEDLYAN---INDI---VDGLAFAGDKMLFYKONDOKKA 203
      743 KIKNDIDTLNQKIDKSIETLF-EIKENSENHIDEIKGQIDKLKVPNKTWF--NEDPKXI 799
      204 EIKI-RMPKIKETSESEPPYSSYGNVLELGGD---LSKNPDLUTKES-GKIY-----254
      800 EKKINIVKIKDKKNIKYKIDKLNEISKLINDKTSLEKKNINLSYKSGISGNIFLQOI 859
      255 -SDSEKQOYLK-----DNI-----ILRKYA 275
      860 DEKKKAHEHTIQAANAVIDDLNIIKKSGOEIEKENIMMDIMDIHKEMKALNISHDYK 919
      276 LKVTI--YVPGKIDMLBNGVYSKEDIATOKANPNLASETTIYADSRNVEDGRSTOS 333
      920 IYHTTSKNHEKISDIRKNSLKIIDFSESYIINDIKKELEKNVLESQNNNTDINO---975
      334 VLSALDQ-FNIRYQVLFKKNKGKALDKGNLVTSBSKLVLFSGDKDEYTG-387
      976 -YLSKIENTYINIL-----KLNKIKKIIDK-----VKEXTDEIEKKN 1010
      388 DKFNYE-----AIKEDGSMFLIDTPKPVNLSDMKVYFNPSKSNKIYVRPEFYLRGK 438
      1011 KINABLNSSEKIITQLENSSLEKCSQK-ISTTIDNVSECIKN---ITVLKTYI---1063
      439 ISDKGGFNWEKLVNESVVDNY-LIYGDHIDNTRDFNITKLVYKDDIDMDGKDYKANGF 497
      1064 VNEKNNINITYFGNAEBYQNVSLNFNINEMADTKQYI-LNIK-----KNNG-1109
      498 PDKVTDMGNVYLGQYSPLAKAVGHYQPLDYNVKPEVINIDPKGNTSIEYADGKSVF 557
      1110 -----TNNTDYNIKELKEHK-----KSNVYDEAGKNTQEIKNKNEI--1147
      558 NINDKRNNGFDGEIOEHYINGKEYTSFNDIKQIIDKTLN-KIIVKDF--ARNTVKE 614
      1148 -----FEKYBQVTVLNL-KTYAV--ELKAKFDKTYNSQOIKIKIKNANTFTSQ 1195
      615 FLINDTGBVSELKPRVYVITONGKEMSSITVSEDFILPVYKGELEKGYQFDGWEISG 674
      1196 -ADSEKKKNEIKNEQRIEDEVAKNKSN-----KALID-----1229
      675 FEGKADAGVYINLSKDTFIKPVFKIEE---KKEENKPTFVSKKKONPV-----723
      1230 -----IOLSTPEF-KIFELKIKIDRTSDCLKTKOIEKISLSDTOETKLI 1278
      724 -NHSQNLNESHKEDLQREHSQKSDSTQVATVLDKNNISKSTTNPNK 773
      1279 EKKNILLTLEKLESUKOKKNIEDOKK-----LDEVANSIKNIIESNVNQ 1324

```

RESULT 8

```

hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18440
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18440
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4550 <LAW>
A:Cross-references: UNIPROT:O77336, EMBL:Z88547, NID:e1325376, PID:e1325396, PIDN:CA8111
A:Map position: 3
A:Map C0425w

```

```

Query Match      5.7%; Score 228; DB 2; Length 4550;
Best Local Similarity 21.6%; Pred. No. 0.11;
Matches 193; Conservative 119; Mismatches 347; Indels 236; Gaps 43;

      5 IAEKPKNLNGKE--GSLKQDTTGV-----EHHQENESIKKKSFTIDNRNISTIRD 56

```

```

Db      746 IEESHSISGSHNEIYASIKCAQNGVRDNKLDGHHNDKDKDNKKD---DKDNDKKD 802
Qy      57 FENKDLKLLIKKKFREVDFTSETGKRMEEYDYKDDKNIIAYDDGTLEYE----- 109
Db      803 GPKKDDKDKKDDKDYDDGDD-----KYDDDNKTDDDDDI---YDDDDPFNDHDSERSK 853
Qy      110 -----TEKLDEIK-SKIYGLVSPSKDGHFELIGKIS-NVSKNAKVTYTG 150
Db      854 RLSNVDLILKDKKKKSGILKKEEDTLKCSKM1NVOEKESSKMYCMATNNCNVYKE-KNYV 912
Qy      151 NNYKSEIKATKYDHFHSTKMTFEDLVANINDIYDGLAFAD-----MRLF-----YKDN 198
Db      913 NDYMH1INNLSHHEMINMKSPG1NNLFKDYEHFHKRDF1NNILKLIKETIIQVLCN 972
Qy      199 DQKKAIEIKRMEPIKETSEYRYVSSYGNV-----TELGEGLS-----KN 240
Db      973 NYKQVCVFCKTTGIR-TKCMFPCSTYFHIYCYNNKQWHRKKNDDQCSQKSKRN 1031
Qy      241 KPDNLTKQE--SGKIYSDSEKQOYLKDN1ILRKGYALKVTTYNPCKTMDLEGNVYSKE 298
Db      1032 QRAASVVRGAEKDKKDSRSLDRKDKYTRRSNKYESANDSDGKN--NGDDKKKD 1089
Qy      299 DIAKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKMDK 358
Db      1090 NIGKNNNNNNNNDDSDSNNNGDSNNNG-----DGNNNINGDNNNNINGDN 1143
Qy      359 EAIDKQGNLVT-----DSSKLVLFSGDKDEYTGEDKFN 392
Db      1144 NNINDDGNNNNCLBEKNGDNGKRRKRRKRRNNNDTSILVLANKDSRQKKNKYG- 1202
Qy      393 EAIKEDGSMFLDTYKPNLSMDKATFNPSPSKNKIYVRNPEYLRGKI-SDKGGFWELRV 451
Db      1203 ---KDHS-----DOEKDPSKINNRSKN-----KVKDKKEVGN---I 1234
Qy      452 NESVDNYLIYGLH1DNRDFN1KLVNMGD---IMDMGMDKYANGFPKPTMDMGN 507
Db      1235 NDDKINN-----NNDNNKNNN1--NGDNKNNNNINGDNKNNNIN--DKI---NKN 1278
Qy      508 VYLQGYSDLNKAVGVHY-QELYDNVKEVNIIDPKGNTSIEYADGKSVFVINDK--- 562
Db      1279 INGDKINNNDK1SKYHSHNYDN--E1IHBQKNPN-KNKTTRKRSINDKEMD 1335
Qy      563 ---RNGEPDGE-----IQEHIYINGKEYTS---FNDIKQI1D1KTLIKIYVQPA 607
Db      1336 ILNNKNNDDK1KLNKFNFLMEYOKIISDKITSGLSNMMDIKNIKD-----IKQIK 1387
Qy      608 RNTYKPEFLNKDTGSEVSELPKPRVTV1IONGKEMSTVISEEDFLPYKKELEKGYOF 667
Db      1388 DIKNIKDI---KQIKDIKDIKQIKQIKVANKIKD-----IKSNED----- 1424
Qy      668 DQMEISGFEKKDAGYV-----INLSKDT--FIKVPFKLIE--KKEENKPTPVSKKK 718
Db      1425 ---IKDIKDVNDVSVHLNKCIVNKKSDTQDFCPSDDKNINN1IKKNKINITDQY--- 1477
Qy      719 DNPQVNHSQLNSHKKEDLOREHSQKSD-STKDVATATVLDKONISSKSTNNPN 772
Db      1478 -NPVDSSTSDCNKQYKKEKAVAKIGMNIKIKLHNNNNNNNNNNNNNNNNNN 1531

RESULT 9
T18429
Hypothetical protein C0345w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18429
R:Lawson, D.; Bowman, S.; Barrell, B.
Submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18429
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1711 <LAW>
A:Cross-references: UNIPROT:O77322; EMBL:Z98547; NID:e1325376; PID:e1325381; PIDN:CAB111

```

```

C:Genetics:
A:Note: PF0345w

Query Match      5.5%; Score 221.5; DB 2; Length 1711;
Beet Local Similarity 19.7%; Pred. No. 0.05; indels 243; Gaps 37;
Matches 160; Conservative 134; Mismatches 274;

Qy      29 EHHQENESIEKESK-----FTIDRNISTIDFENKDLK-----KL1KKKFRVD 74
Db      984 DNNNDNNNNMYDSSSSSNHNYI1ITNDKRLN-MDNF1NNNL1EINNQNVL1EKL1EYIN 1042
Qy      75 DF-TSETGKRMEEYDYKDDKNIIAYDDGTLEYETEKLBDEIKSIYGLVSPSKDGHF 133
Db      1043 NVKLTSTS-----NYESGNTNSKDE-----HNISDSKSKEDDTLNSKSSYE 1086
Qy      134 ILGKI--SNVSKNAKVYVYKSEIKATKYDFHSTKMTFEDLVANINDIYDGLAFAGM 191
Db      1087 YNNK1LQSTSNKSLGAYENNLFSKKKGNK-----GYLKD1-EHINDIQD--KYPEDL 1138
Qy      192 RL-----FVKNDQKKAIEIKRMEPIKETSEYRYVSS---YGNVIELGEGLSKNRPD 243
Db      1139 NINCYNKYVIEBEKHL-----LPLEL-----EYVLVSDEKFG-----LNKIKND 1179
Qy      244 NLTKMESKIIYSDSEKQOYLKDN1ILRKGYALKVTTYNPCKTMDLEGNVYSKEDIKI 303
Db      1180 N-----NIYMKHQNYHNLVDN--OKKHILFDTNKN-----VSI 1212
Qy      304 OKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKMDK--GEAI 361
Db      1213 QR-NNNINSV1KTNH1EYERKNNKQDN-----YDNFCDDKKKIYVYII 1255
Qy      362 DKDGNLVYDSSKLVLFSGDKDEYTG-----EDKFNVEAIKEDGSMFLDTYKPNV-- 411
Db      1256 NSDKDIYHNN--I1YTKNEKEG1GNH1NRNDKQITFELK1DGVKPEFDTFKDSYID 1312
Qy      412 --SDKATFNPSPSKNKIYVRNPEYLRGKISDKGFNMELVNVEVDNYLIYGLH1DN 469
Db      1313 CHNKKEN1LNTNNKK-----EDHQLIDVADK-1FN--ETNM1TYMNNK1IYDKVNH 1362
Qy      470 TRDFN1KLVNMGD1MDMGMDKYANGFPDPTMD--GNYVLQGYSDLNKAVGVHYQ 527
Db      1363 KCKTNDV1HNNMD1LS1KNNESYHKLIDES1LSNKKNNKVS1L-----L1NNNKDSS 1505
Qy      528 FLYDNVKEVNIIDPKGNTSIEYADGKSVFVINDKRNNG-----FDGEIQEHIYINGKEY 583
Db      1408 ---IN1QEDDDDDHNN-----NHNNNNNNNK1LFEYTKQDQ1LHNNKNNL 1452
Qy      584 TSPNDIKQI1D1KTLIKIYVQDPARNTTVKPEFLNKDTGSEVSELPKPRVTV1IONGKEMS 643
Db      1453 EGTBEFSDPFIERKNNKIK1KNNESYHKLIDES1LSNKKNNKVS1L-----L1NNNKDSS 1505
Qy      644 STVISEEDFLPYVYKKELEKGYOFQDQMEISGFEKKDAGYV1NLSKDTFIRKVPFKLIEK 703
Db      1506 SY-----DNNKNNNNKNNNNK-----NANNEN 1527
Qy      704 KEEENKPTFVYSKKKNDPQVNHSQLNSHKKED-----LOREHSQKSDSTKDVATV 756
Db      1528 NNNKNNNNNDSPFSKDNNL1NNNDNNNNNNNDSPFSKDNNL1NNNDNNNNNNNNKVIKKEI 1587
Qy      757 LD--KNNISSKST-----TNPN 772
Db      1588 IDDEKND1HRRDNIY1KDVSVSPL1NNPN 1618

RESULT 10
A24594
Probable major surface antigen (83k, 19k, 42k) precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jul-2004
C:Accession: A24594
R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholas,
Nature 317, 270-273, 1985
A:Title: Primary structure of the precursor to the three major surface antigens of Plas

```


Db 812 DPLDLFNQNNIPWYMSFSLNLSLSOLFMEIYEKENVCNLKYLKONDKIKLLEBAK 871
Qy 598 NIKIVKQFARNTTVAEFLINKDTGSEVSELKPRVVTIIONGKEMSTIVSEEDFLPY 657
Db 812 KVTSTVKTLSSSSMQPLSLTPQDKPEVSANDTSHSTNLNLSKLLENILS----- 922
Qy 688 KELEKGYFQDGEWISGFEK-----KDAQVYNLSKDTPIKPFVKKEIEKKEENK 709
Db 923 LKKNKVIYQ---ELIGQKSSNFYEKILKSDTFFNESFTFVKSADINDLSDESKR 978
Qy 710 PTF--DVSKKKDNPOVHNSQLNESHKEDLOREBHSQKSDSTKDTATATVLDKNNISSK-S 766
Db 979 KLEEDINKTKTQLSLFPLYNKYKLERLPDKKTVKRYMOIKKLTLLKEQLESKLN 1038
Qy 767 TTNMP 771
Db 1039 SLNMP 1043
RESULT 12
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
N.Alternate names: gp195 surface antigen
C.Species: Plasmodium falciparum
C.Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C.Accession: S05603; S04850
R.Myler, P.J.
submitted to the EMBL Data Library, April 1989
A.Reference number: S05603
A.Accession: S05603
A.Molecule type: mRNA
A.Residues: 1-1639 <MYL>
A.Cross-references: UNIPROT:P04933; EMBL:X15063; NID:G9896; PIDN:CAA33163.1; PID:G9897
R.Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
A.Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasm
A.Reference number: S04850; MUID:89345116; PMID:2668887
A.Accession: S04850
A.Molecule type: mRNA
A.Residues: 1504-1639 <MYL2>
A.Cross-references: EMBL:X15063
C.Superfamily: major merozoite surface antigen
C.Keywords: glycoprotein, merozoite, surface antigen
F1-15/Domain: signal sequence #status predicted <SIG>
F1-20-1639/Product: major merozoite surface antigen #status predicted <MAT>
Query Match 5.4%; Score 218; DB 2; Length 1639;
Best Local Similarity 18.6%; Pred. No. 0.079;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;
Qy 23 KDTTG-VEHHQENESIK-----EKSFTIDRNISTIRFENDKDKLIRKGFREYVD 75
Db 232 KDNVGMEDYIKNNKKTINELNLEBSKKTIDKNNAKREB-----KKKLYQAQY 283
Qy 76 FTSEFGKMEYDYKDDKGNIAVDGDTLVEYTEKDEISKIYGVLSPKDGHFEL 135
Db 284 DLSTYKKOLEE-----AHNLISV-----LR--KRIDTLK-----XENIKKEL 319
Qy 136 GKISNVSKAKVYVNNYKSEIKATKYDFH-----SKTWFPDIYANINDIVOLAVA 188
Db 320 DKINEIKNPPRANSNTPTPLDKNKKIIEHEKEIKELAKTKIKFNIDSLFTPL----- 373
Qy 189 GMRLLFVNDONDKAKIIRMEKIKETYSBP-----YVSSYGVN-----IELGE-GDLSK 239
Db 374 -ELEYYLRKKN-KNIDISAKVETKSTEBENYVNGVTYPLSTNDINNALNELNSFGDL- 430
Qy 240 NKPDLUTLKESGKISDSKQOYL-LKNNIIL-----RKGAALKATTYNNPGKTYML 289
Db 431 -NPFYTKSPSKNIYTDNERKKFINEIKIKIEKKKIESDKSYEDRSKIN-----DIT 485
Qy 290 EG-----NGVSGKE-----DIAIKOKA-----NPNLRALSETTYVADSNVDEGSTOSV 334
Db 466 KEYEKLMEIYTSKPRNNIDLTNFEKMGKRYSYKVEKLTTHNTPTASYSNGKINLEKLIK 545

Qy 335 LMSALDGF---NI-----RYQVFTFRQNDKGEAIDKGNLVYDSSKLVLFQKDKKEYT 385
Db 546 ALKVMEDYSLRNITVEKEIKKYKNLISKLENIETLVE--NIKROEQLF-----EKKIT 598
Qy 366 GEDKFNVAIKEDSGMLPTIDKFPVNLSDMKYFNPSSKNKIYVRNPEFYLAGKISDKRGF 445
Db 599 KDNKPRDKILIEVSDIVAVQOQVYL---NMKIDELKKTQILKNVE--LKINIHPVNSY 653
Qy 446 NWELRVNESVVDNYLYGDHLIDNTRDFNIK-----NVK----- 480
Db 654 KOENKQF-----PYTILVKKELDKLVFMPVAVESLINEKKNIKTREGSDNSEPTBEGI 709
Qy 481 -----DGDIDMGMKDYK-----ANGPDKV 501
Db 710 TGOATTKPQOAGSALBEDSVQAOQOQKQAPVPVPVPEAKAQVPPAPVNNKTENV 769
Qy 502 TDMGNVYLGQYSQNLN-----KAVGVHYQFLVDNVPVNNIDPKMTSIEYDQKGVV 556
Db 770 SKLD---VLEKLYERFLNTSYCHKYLIVSHSTNMEKILKQYKTYEBSKLSSCDPLDL 826
Qy 557 FNIN-----DKRNGFDG---EIQEQHIYNGKEYTSFNDIKOIIDKTINKIIV 603
Db 827 FNIQNNIPWYMSFPLSLNLSLSOLFMEIYEKENVCNLKLDONDKIKLLEBAKGVTSV 886
Qy 604 KDFARNTTVEFLINKDTGSEVSELKPRVVTIIONGKEMSTIVSEEDFLPVYKGELEK 663
Db 887 KTLSSSMQPLSLTPQDKPEVSANDTSHSTNLNLSKLLENILS-----LGKKN 937
Qy 664 GYQPDGWEISGFEK-----KDAQVYNLSKDTPIKPPVKKEIEKKEENKPTF--D 713
Db 938 IYQ---ELIGQKSSNFYEKILKSDTFFNESFTFVKSADINDLSNDSSKRKLEED 993
Qy 714 VSKKKDNPOVHNSQLNESHKEDLOREBHSQKSDSTKDTATATVLDKNNISSK-STNMP 771
Db 994 INKTKTQLSGDLYNKKLKLRLPDKKTVKRYMOIKKLTLLKEQLESKLNINMP 1052
RESULT 13
G90603
lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C.Species: Mycoplasma pulmonis
C.Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C.Accession: G90603
R.Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galieson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A.Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A.Reference number: A9512; MUID:21267165; PMID:11353084
A.Accession: G90603
A.Status: Preliminary
A.Molecule type: DNA
A.Residues: 1-1546 <KTR>
A.Cross-references: UNIPROT:O98P17; GB:AL445566; PID:G14090150; PIDN:CAC13908.1; GSPDB:G
A.Experimental source: strain UAB CTIP
A.Genetic code: SGC3
Query Match 5.4%; Score 217.5; DB 2; Length 1546;
Best Local Similarity 20.2%; Pred. No. 0.077;
Matches 180; Conservative 125; Mismatches 336; Indels 249; Gaps 40;
Qy 1 KLGIAESKRFNLTNGKSGSLK-----KDTTVEHHQENESIKESFTIDRNISTIRD 56
Db 103 KSKELAESQVGVSVNNEKNRFLLEAKTKTKTKONPSAKSQNLDTSGDHLNKSQKSNISN 162
Qy 57 FENKDLKLIKKEPREVDFTSEFGKMEYDYKTD-DKGNIAVDGDTLVEYTEKDE 115
Db 163 IKNEQSKTLQTLNNEEHSANLQTONLILESNTQODLDPN-----NSNGKEKEKLN 214
Qy 116 IKSXIYGVLSPSKQGHFILLK-ISNVSKNKKVYVNNYKSEIKATKYDDFSKMTPTDL 174
Db 215 VLS--FNKQEAIVKQDSFEFFYFREIVSKLESKINVDKRVDEITLDTKTRADF-----SFEQ 267


```

QY 175 YANINDIVDGLAFAGMRLFLVKONDQKAEIKIRMEPIKETKSEYPRVSSYGNVTELG 234
DB 268 Y--IEKIDKLFKMAST-----IKDTYQ-----TNKIFLTNDEY-----238
QY 235 GDSKRKPNLTGME-----SGKIYS-----DSEKOYLLKDNITLIRKGYAL 276
DB 299 --IKKNESNLKDKKMFNNFENIYYFLEKDLALVEIOALHKQYLERKEIITKGEIYD 356
QY 277 KATTYVPGKTDMEGNG-VYSKEDIKAI-----OKANPNLRALS 314
DB 357 KAAAFIKSRKELSKAGLFLSKQSAKITQOLINHIIISRPPEAPENLELSKAMGLINEL 416
QY 315 ETTIVADSRNVEDGRST-QSVLSALDGFENIIRYOVFTPKMDKGEALIDKGNLVTDSSK 373
DB 417 KPEIYVEKEGVEIAYPTLQDAISNAQDGQKIFL-----NKNLKLDR--SIVVDKN- 464
QY 374 LVLFQDDKEYTEED--KRVNEAIKEDGSMFLPIDTPKVNLSMDKATFNPS-----421
DB 465 IITFAKSNVLTITKDKSSKSPFTMFIYOKGALTFEIAEPSSQSINLNGIGTSFKDESSILVKI 524
QY 422 -KSNKIYVNPBEFLYLGKISDKGFMWELRVNESVVDNYLYIGDLHIDNTRDN-----474
DB 525 EKRAKLVAKTGTAFINSKFSKYG-----SVFEN--YGSVVIKGAKTNNVNSSEG 572
QY 475 -----IKLVKDDIDMDGMKDYKANGFPDK-VTDMGNYVLTQVSYDLNAKAVGYH 525
DB 573 GIIRNVGSSSLTFPKNEI-----RDNISTG--DKGIISGNTAISGGSIDGKS-----620
QY 526 YQFLYVNVKPEVNIDPKNTSIEYADGKSVENINDR-----NNGPDGEIQEOHIYING- 580
DB 621 FRSLLINLE-KTININSGSIVNNAASVKSILFEIDNSKIQISNNMALINPFGSAIFLAKN 679
QY 581 -----KEYTSFNDIKOIIDKTLINKIV-----602
DB 680 STMHLAGSLKIKKEASBOQIEVLVDLPQAKLISPKNITSLDNYOKLSAIFKIPSVKN 739
QY 603 VQDFAR-----NTTVEFLINKDT-----GEVSELKPHRVVTIIONGKMSSTIVSE-DF 652
DB 740 INDFKVPLVWNTKEKFFKLPMPDTKLFPVNFYKTLKQNHDLIQSGPESFEKIIIXDELDF 799
QY 653 -----ILPVYKGELEKGYQF--DCWEISGFEKSK-----DAGYVINTLSK 689
DB 800 YRPPTAAVKKLLITQLVRTIPKTHKFWMAFEYENPFLQKWEYETINELIRIDPFYLDIAY 859
QY 690 DPFIRKVFKEKKEENKPTFDVSKKONPQVNSQLNESHREKDLQR 739
DB 860 PEFV-----ENGKMLPRPEY-IHTNVVNPVLEH-----PRNEDVAR 894

```

RESULT 14

reticulocyte-binding protein 2 - Plasmodium vivax (fragment)
 C1:Species: Plasmodium vivax
 C1:Date: 12-Mar-1993 #sequence_rev1sion 12-Mar-1993 #text_change 09-Sep-1997
 C1:Accession: B42771
 R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
 A1:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
 A1:Reference number: A42771; PMID:92315338; PMID:1617731
 A1:Accession: B42771
 A1:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A1:Molecule type: DNA
 A1:Residues: 1-1252 <GAL>
 A1:Cross-references: GB:M88098, NID:g160627, PID:g160628
 A1:Experimental source: strain Belem, merozoites
 C1:Genetics:
 A1:Gene: RBP2

Query Match 5.4%; Score 215.5; DB 2; Length 1252;
 Best Local Similarity 20.9%; Pred. No. 0.071;
 Matches 207; Conservative 130; Mismatches 324; Indels 331; Gaps 50;

```

QY 2 LGELIAESKPK---NLNGKESGLKDTGTGVEHHQDNEESIKESFT-IDRNI-----51
DB 266 LSEIKKQKCTTEIENSRRG---KDKLEFLKFKPNESENKNNKININENINRSEQY 322
QY 52 -----STINDFENKDKLILKKA--PREVDFFS 78
DB 323 LKDIEDAEKQASTKVELFKRHBTTSNI FKESBIIIGVEFKSQKINKAKADIMEKELERHNS 382
QY 79 ETKGMEBYDYKDDKGNIIAYDDGTDELEYETEKDEISKITYGVLSPKQDGFELTIGKI 138
DB 383 ELQYQVKGQFQENLANKLNPENTDMAED-ELANDKSNKAVLIETNLESYKHN-----LSEI 437
QY 139 SNVSKNAKYLYGNNYKSI--EIKATKYDFHSKTM-----TFDLVYAN-INDIVDGLAFACD 190
DB 438 TYNKQGEKIY--SKADIMQIKATSENPAETKLEKVKDQSNVYVNLQITF-----E 490
QY 191 MRLFYQDND-----QKAEIKIRMEPIKE-----TK-----S 218
DB 491 RNLIVTEKRLNGIDSTTINIGALKEKSGNVEIGLEKLEIIGNRKLVKVDITYKKSINS 550
QY 219 EYPVYSSYGNVTELGEDLSKKNPDLTGMESKITYSDSEKQOYLLKONILIRKGYALV 278
DB 551 TYGNFSSSLFNNPDLNDQYDPNKNINDYENKM--GEIYNFEGSLNKISENL--RANSEMT 605
QY 279 TYVNPCKTDMLEGNVYSKEDI---AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVL 335
DB 606 SDYNSAKTLRLBA-----QKEKYNLNLKEEANKYLR-----DYKVESPR-----646
QY 336 MGALDGFNIIRYOVFTPKMDKGEALIDKGNLVTDSSKLYLFGKDDKEYTGEDKRVNAI 395
DB 647 -----FIFNNK--ESLIDKINEMIKKEQLTV-----NEGHGVNKKQLENI 683
QY 396 KEDGSMFLPIDTPKVNLS-----MDKN-----YNPERSKMKIYVRNBEFLYRGKI 439
DB 684 KE-----LVDEN--NLSDLKQATGKNBEIQKITHSTLKNKAKTILIGHDTSAKTYGICI 736
QY 440 SDKGFNMELRVNESVVD-----NYLYIGDLHIDNTRDNIRKLVNVD-----481
DB 737 TP-----ELALTELLGDAKAKTQAQELKPKSKNNVVLLETNNNSKQNTNELDVHKNIQDAYK 790
QY 482 -----GDINDMKQDYKANGFPDKVTDMGNYVL-----QTVSYDLNAKAVGYH 525
DB 791 VALEILLASHDEIDTQKD-----SSKLIEMGNQIYLVKVVLIQYKKNKISSIKSEEAVS 844
QY 526 YQFLYDNYK-----PEVNIDPKNNTSIEYADGKSVENI-----NDKRNNGPDEIQOQ 574
DB 845 VKI--GNVSKKSELSKITSYDNITIALKQTELQNLNNSFTQEKNTNNSDKLE-- 900
QY 575 HIYNGKEYTSFNDIKOIIDKTLINKI-----VVVD-----605
DB 901 -----KITDFESLKNML-KTLBEVGNALKASDNHHEVQSKSPVPVNPALSEIEKEED 953
QY 606 -PARNTTVEFLINKDTGVESELKPHRVVT--IONGKMSSTIVSEDFILPVYKGELE 662
DB 954 IOLSWTALDELKCKRTCEVSHYKLIKQVTKEISDDTELIANTIKEN--YKAYIAYIK 1009
QY 663 KQYQDGMKISGFEKQKAGVYNLSKQTFIPV-----PKIKBEKKEENKPTFD-- 713
DB 1010 KNYE-----DVIQD--VLTINHEPNTQVSNHNEPTNDKNSKSEELTKAVTDSK 1057
QY 714 --VSKKK-----DNPQVN-----HSQLNESHR-KEDLOREBHSQ 744
DB 1058 TIISLKLKGVIIENVNENTENNTIESSAKETALAYNELKAKKTSILNEIYQSNBEVKAQEMKS 1117
QY 745 KSDSTKDVTA--TYLD--KNNISSKSTTN 770
DB 1118 NADKRYDVSQKIPNTVLDIQKSNIVTNQHSINN 1149

```

RESULT 15

A45597
 DNA-directed RNA polymerase (EC 2.7.7.6) III largest chain - malaria parasite (Plasmodi
 C1:Species: Plasmodium falciparum

C|Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jun-2000
C|Accession: A45597
R|L1, W.B.; Bzik, D.U.; Tanaka, M.; Gu, H.M.; Fox, B.A.; Inselburg, J.
Mol. Biochem. Parasitol. 46, 229-239, 1991
A|Title: Characterization of the gene encoding the largest subunit of Plasmodium falciparum
A|Reference number: A45597; MUID:92018020; PMID:1656254
A|Accession: A45597
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-2339 <L1>
A|Note: Sequence extracted from NCBI backbone (NCBIN:61099, NCBIPI:61101)
C|Superfamily: Plasmodium DNA-directed RNA polymerase II large chain
C|Keywords: nucleotidyltransferase; transcription

Query Match	5.3%;	Score 214;	DB 2;	Length 2339;
Best Local Similarity	21.6%;	Pred. No. 0.18;		
Matches 191;	Conservative 134;	Mismatches 306;	Indels 254;	Gaps 50;

OY	24	D7YGEHHNOH--NSESJIKESSFIDBNISTIRF--ENKDLKJLKKKERFREDPTS	78
Db	1292	DNTYVEQJEMELSSGNKTKKEQSKF----GTIRMHEDSEGNKFTTKAK----	PFI 1342
OY	79	ETGKMEEBYDKYDDKGNIIAYDOGTLEYTEKLEDKISKIY-----	-G 122
Db	1343	EK-KKGRKHECNDIDIEYNNTOYD--NIYANNISCNKYKSQNLENTHHQVANNDSLFIKAN	1398
OY	123	VLSPSKODH-----PEILGKISNVSNAKVYGGNNYKSIEIKATUYDHSKTMFDL	174
Db	1399	VILPREKXHSIFHPVNDYRNVAVELKNLMDKKKIFLANSNKA--VVOSSKYNMSSKNLKRTI	1456
OY	175	YANINDIYDG-----LAPAGDMRLFYKXNDQ--KKAIEIKIRMEK----	IKETKSE 219
Db	1457	EI-INNIIRNEKKKLNLRKTKTMDNDNDYWSDDDSIIAKKIILKIKNEKRAKHYHREKEN	1515
OY	220	YPYVSSYGNVTELGEGLSKNKPDPNLTK-MESGKYSDESEQOYLKDNITILRGVALYK	278
Db	1516	FDR--NNYKMITDNNNDNNNNNDNNNNNDNNNNNNNNNNNNNNNNNNNYUHLDDV----	NNIGV 1569
OY	279	TTYNPGKIDML--BENGYSKEDLAKIOKANPUL-----RALSBTTIYADS-	322
Db	1570	TNYN-----TNYIPNDONGIYEKXTNNNELTTNNSMCDKNNDPSDEFENNINENDLLYDXYK	1626
OY	323	-----RVNEDRSSTQSVLMSALDGFNIILRYGV--TFKANDKGBALDKDNLVTDSGL	374
Db	1627	YRQIFKQVITGVSVSEYVESYKQHIILRPYELIKMTSFLSLYELIIPNIFLHTKJLQSK-	1685
OY	375	VLFPGKD--DXYTGEDKFNVBAIKEDSGMTFIDTKPVNLSMDKNYFNPSKSNKIYARNP	431
Db	1686	-----KEKPTHQKNGKMKIYIEIKK--WLF--KAINI--XCYFSEFKSIEI----	1728
OY	432	EFLYLRGKISDGKGFWMELRVNSESVDNVTLLIYGDLDHIDTRD--FNI-----	KLANYKD 481
Db	1729	-----IKKKDYFNYYIIT-KYDISHRYIHH-DYSPNLKQULYFIFENIYKFFYISTP	1779
OY	482	GDINDMGKDYKANFPDPKVTMDGNVYLQGTGSDPLNAKAVGNHGYOFLYDWNKBEVNID	541
Db	1780	GD-----AVGTSIAQSIGEPGTQMTLKTGHPAGVASMNV-TLGV-----	PRIKELIASN 1828
OY	542	KGNTSIEFADGKSVFVINDKRNNGF-----DGEIOEHITNYINGKYYTS-----	F 586
Db	1829	SIQFPI-----LNTPELVNDVNNFALMWKSKLEKTIITJIDICMYIK-EDYTSRGVFLSVKF	1882
OY	587	ND-----IKOIIDTKLNI-----KI--VWQOPANTVVK--EPLINKOT	621
Db	1883	NBELIQKFLYNANVNIKOIILKQSHINKIKINIKIHINIVINKYLAHLSKNDERIF----	1938
OY	622	GEVSELKPHRYVVTIÖNGKEMSSSTVSEEDFLLPVYKGELEKGYOPDGMSEISGFEKG--K	679
Db	1939	FQWESLKKGGLDLILYGDOKIRCIKKEDI-----EYDNEDEICD	1980
OY	680	DAGYVINSKOTFIKPFVKKIEEK--EENKPFPIVDSKKNDPQVNHQSLNESHRED	736
Db	1981	DMDEYVNSQGI-----ELYERKCSNEKKNKARVYKEIID-----DN	2019

```

Qy      737  LQRE-----HSQKSDSTKDVATVLDKNISSKSTNN 770
          ||::||      :||:|      ||:|      :||
Db      2020  LEKENIIYVSEKDSVNLKSEKKDIN----DDNNNNDDNNNNN 20600

```

Search completed: February 16, 2005, 19:21:34
Job time : 51 secs